Psychometrika

CONTENTS

VOLUME THREE SEPTEMBER 1938 NUMBER THREE

PSYCHOMETRIKA—VOL. 3, NO. 3
SEPTEMBER, 1938

A THEORY OF LEARNING AND TRANSFER: I*

HAROLD GULLIKSEN AND DAEL L. WOLFLE
The University of Chicago

A rational theory of discrimination learning is developed for the special case in which the subject must discriminate between two stimuli which differ with respect to one variable such as size or brightness. It is shown that the previous equations developed by Gulliksen and Thurstone are special cases of the present one. It is predicted that the ultimate level of accuracy of the discrimination is inversely related to the difference, as determined psychophysically, between the two stimuli. Other implications of the theory for experimental work are presented.

INTRODUCTION

This paper will develop a rational theory which will describe the course of learning of a sensory discrimination; predict the relative difficulty of different types of discrimination; and predict the nature and accuracy of transfer of the learned response to new stimuli. The prediction of both learning and transfer data from the same postulates indicates an order and unity underlying these two processes which has not previously been demonstrated.

The theory is considerably more general than the previous developments of Thurstone (28, 29) and Gulliksen (4) which were confined to a description of the course of learning. These earlier theories will be shown to be special cases of the present one.

The present theory unites the three fields of learning, transfer, and psychophysics, in that it utilizes information obtained from psychophysics regarding the psychological similarity and dissimilarity of stimuli to assist in predicting the difficulty of learning and transfer experiments involving those stimuli.

The nature of the relationship between psychophysics and learning may be stated briefly as follows. Other things being equal, if two stimuli are very much alike, as determined by psychophysical methods, it will be more difficult to learn to respond differentially to them than if they are very different. The phrase "other things being equal" takes care of such variables as the learning ability of the animal and its initial attainment.

* We are grateful to the members of Professor Thurstone's Seminar in Mathematical Psychology for criticism of this paper and particularly to Mr. John Reiner for assistance in the derivations involved.

The relationship between psychophysics and transfer may be similarly stated. Other things being equal, the more alike two stimuli are, the greater is the probability that a response learned to one will transfer to the other.

These two statements imply that an inverse relationship exists between learning and transfer experiments. Where learning to discriminate between two stimuli is difficult, transfer is easy, and *vice versa.*

It will be shown that the theory leads to the logical deduction of a number of experimentally known facts regarding discrimination. Further, it predicts a number of facts and relationships which have never been tested experimentally. Experimental tests of these deductions will provide a crucial test of the adequacy of the theory. Until the psychophysical experiments necessary to determine the degree of similarity or dissimilarity of the stimuli used have been performed it will be impossible to predict the quantitative results of any single experiment. At present the theory is limited to the prediction of certain *relationships* to be found between experimental results in psychophysics, learning, and transfer.

In order to explain the development of the theory we will follow through an analysis of the problems of visual discrimination in the white rat. The theory is by no means limited to visual discrimination data, but the development of the theory will perhaps be clearer if it is explained consistently with one type of problem as an example.

The theory will be applied to the usual type of discrimination experiment, using apparatus such as that devised by Yerkes and Watson (34), Lashley (18), Fields (2), or Munn (21), in which the animal receives a food reward for selecting the correct one of two or more similar openings. Before beginning the experiment proper, the animal is taught to find food in the apparatus. When formal training is started, the openings are closed with doors which are marked with the stimuli to be discriminated. The stimuli usually consist of geometric figures which differ in size, color, form, or brightness. Whenever the animal approaches the door bearing the positive or correct stimulus it can get through to receive a food reward. Whenever it approaches the door bearing the negative stimulus it can not get through, and may receive some form of punishment. The two stimuli are presented in the right-left and the left-right order at random to prevent the animal from developing a position habit. The discrimination is said to have been learned when some arbitrary criterion of correctness is reached.

From the viewpoint of the present theoretical development, the essential features of such experiments may be listed as follows:

1. The response or responses involved have already been learned by the subject so that the acquisition of a motor skill is not a feature of the experiment.

2. The subject is presented simultaneously with two (or more) stimuli, for example, a bright and a dim light or a large and a small circle.

3. The subject is rewarded for making a certain response to the situation such as choosing the brighter light or jumping to the smaller circle; and punished, or not rewarded, for making other responses such as choosing the dimmer light or jumping to the larger circle.

Definition of Stimulus

Various suggestions have been made regarding the aspect of the total situation to which the animal responds. Lashley (19) has recently discussed the several possibilities of the basis of the animal's response. The two most clearly contrasted hypotheses are: (1) that the animal responds positively to one of the stimuli or negatively to the other; and (2) that it responds to the relationship between the stimuli, *i.e.*, greater than, dimmer than, etc. The first of these definitions implies a response to the absolute characteristics of the situation, and the second implies a response to the relative characteristics. In contrast to both of these definitions, we suggest that *the animal responds to the total stimulus configuration* consisting of two stimuli presented simultaneously in a given spatial order. The theory here presented will be based on this definition. This definition does not imply that the animal must see all parts of the configuration with equal clarity at any one time. The animal may scan the configuration or examine it part by part. The response, however, is, in terms of the present definition, made to the total configuration whether that total be seen at once or built up as a construct of successive examinations of the parts.

To avoid confusion of terms, we will use the word "stimulus" in the conventional sense to refer to the individual lights, colors, or other stimuli. We will use the word "configuration" to mean the pair of stimuli in a given spatial order. Two configurations made up of the same two stimuli, but in the opposite orders, are shown as *a* and *b* in figure 1. The statements about stimuli and configurations apply to such variables as size, brightness, hue, and saturation. Size is used in the illustrations merely because it is easiest to represent graphically.

The conventional account of the discrimination experiment describes the animal as choosing between two *simultaneously presented stimuli*; our account will describe the animal as responding differently to two *successively presented configurations*.

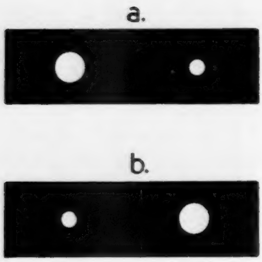


FIGURE 1
Two configurations made up of the same two stimuli but in opposite orders.

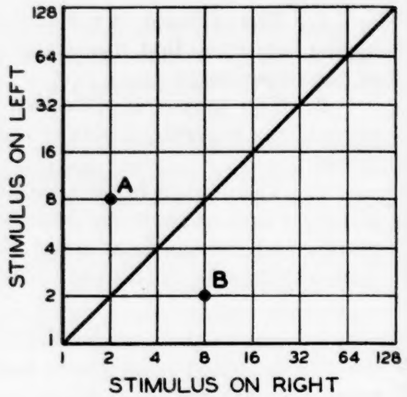


FIGURE 2
Plot of the two configurations shown in figure 1. This plot provides a first approximation to the problem of scaling the configurations.

Psychological Scaling of Configurations

All configurations which might be used in an ordinary two-choice size discrimination problem may be represented graphically, as follows: Represent the size of the stimulus which appears at the right by the horizontal axis and the size of the stimulus at the left by the vertical axis; then point A on figure 2 represents configuration *a* of figure 1 and point B of figure 2 represents configuration *b*. This method of plotting is not limited to size discrimination but may be used to represent the configurations used in any two-choice problem requiring discrimination with respect to one variable such as brightness, pitch, loudness, saturation, etc. In general, any configuration composed of two stimuli differing with respect to one variable may be represented as a point on a two-dimensional plot. Conversely, any point on such a plot represents a configuration composed of two stimuli. On such a graph, the two configurations consisting of two stimuli used in reverse order would be indicated by two points symmetrically placed with respect to a 45° line through the origin.

Such a graph suggests the possibility of a closer relationship between learning theory and psychophysics. The degree of similarity of two configurations may be regarded as a psychological "distance." The more similar two configurations are, the less is the "distance" between them; the greater the difference between two configurations, the greater is the "distance" between them. If a psychophysical meth-

od of determining the distance between two configurations can be worked out, the resulting scale would represent psychological distances between the configurations involved. Such a method, if generalized to the n-dimensional case, would be applicable to configurations involving many stimuli and also to stimuli of any type, including, for example, complex form discrimination. Young and Householder (35) have recently described a matrix method of determining the dimensionality of a set of points where the distances between the points are known. Scaling the configurations, while necessary in order to measure the distances accurately, is not necessary for a general discussion of the relationships involved.

Since the Fechnerian principle of a logarithmic relationship between stimuli and sensations is well recognized, we have adopted it as a first approximation. Hence, in figure 2 the configurations are plotted on the basis of a logarithmic principle.

Definition of Response

The response made by a rat in a discrimination experiment would usually be defined in terms of the relationship to which the animal was being trained to respond. This definition follows from the belief that the most important aspect of the stimulus situation is the relationship between the two stimuli used. Since we have elected to use a configurational definition of the stimulus, it is necessary to change the definition of response.

Révész (24) and others have developed a technique for determining which of several aspects of the stimulus, such as form or color is dominant. By an extension of this technique it would be possible to define the response for any species in terms of the dominant aspect of the stimulus. The ease with which rats learn maze habits and the difficulty of eliminating position habits in training rats on discrimination problems suggests that directional habits may be the dominant type of response in this species. On the basis of such evidence, and experimental data to be presented later, the rat's response may be defined directionally: *The rat learns to jump to the right when presented with one configuration, and to jump to the left when presented with the other configuration.* Our definition of the rat's response in a discrimination experiment is here stated quite arbitrarily. Later it will be demonstrated that it is possible, on the basis of this definition, to predict such diverse facts as the transfer to new configurations on a relational basis and the greater difficulty of learning an absolute than a relative discrimination. The prediction of experimentally known facts on the basis of a directional definition of response provides additional evidence for the value of such a definition.

It is also possible to test the definition directly by experiment. Since the training prior to the transfer tests normally eliminates the possibility that the animal will respond to the test situations on a positional basis, it is impossible to determine from such experiments whether the original learning should be regarded as a relative brightness habit or as two directional habits. A crucial experiment must allow the animal to transfer on either basis. Such an experiment would present the subject with the same configuration (the same two stimuli in the same order) on every trial. The animal would then learn to go, for example, to the brighter light on the right-hand side. After the attainment of a fairly high level of accuracy, new configurations could be presented and the subject rewarded no matter how it responded. The definition of the response would then depend upon the outcome of these transfer tests. If the subject went to the right on the test trials, we would define the response as being a jump to the right. If it went to the brighter of the two stimuli, the response would be defined as being the brighter side. To distinguish between these two alternatives, we performed the following experiment.

The subjects were 24 female hooded rats from the colony maintained by the Department of Psychology at the University of Chicago. The rats were trained on the Lashley jumping apparatus, in which the animal is confronted with two similar apertures in each of which a card can be placed. The cards bear the stimuli to be discriminated. The rat jumps from a small platform some 9 inches from the cards. If it jumps against the correct card, the card falls backward and the animal lands on a platform where it finds a food reward. A block behind the wrong card holds it securely in place so that if the rat jumps against the wrong card it falls into a cloth net some 20 inches below the cards and jumping platform. The fall constitutes punishment for an incorrect choice. Training was continued until the rat attained a criterion of 50 consecutive errorless trials (10 trials per day) in response to the same configuration (the same two stimuli in the same spatial arrangement). It was then tested on a series of 16 different configurations. The stimulus cards were black with white circles on them. Five sizes of circles were used. The areas of the circles increased in the geometric ratio 1:3:9:27:81; the actual areas were 0.26, 0.78, 2.35, 7.06, and 21.15 square inches. For training trials a circle of 0.78 sq. in. was paired with one of 7.06 sq. in. The test configurations consisted of 16 of the 25 possible pairings of the five circles. The training and test configurations are shown in figure 3 and plotted in figure 4.

We followed the customary procedure by first training the animals to jump through open windows with no stimulus cards in them. When this problem was mastered, a white card was placed in one win-

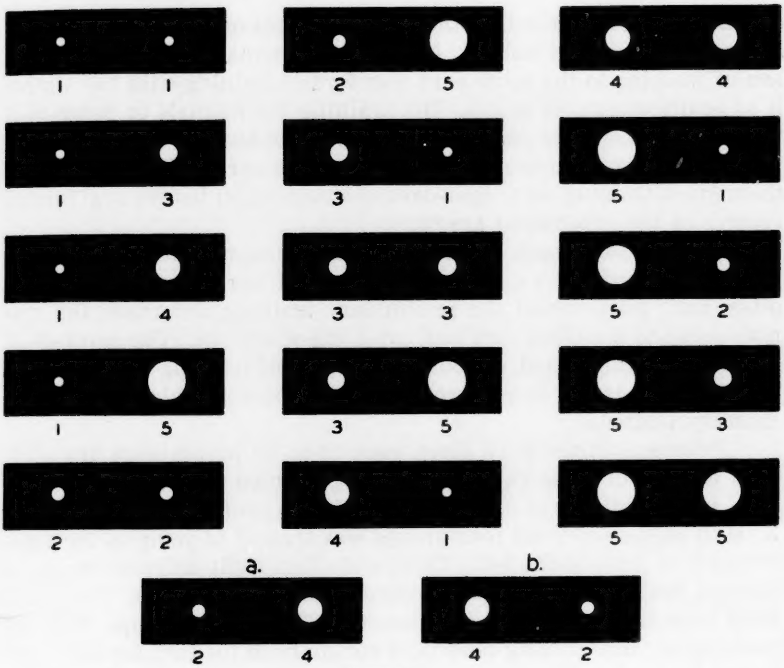


FIGURE 3

The test and training configurations.

One-half of the animals were trained on configuration *a*; the other half on configuration *b*. Each animal was tested on the other 16 configurations shown. The numbers under each configuration refer to the sizes of the circles; 1 being the smallest and 5 the largest of the five sizes used.

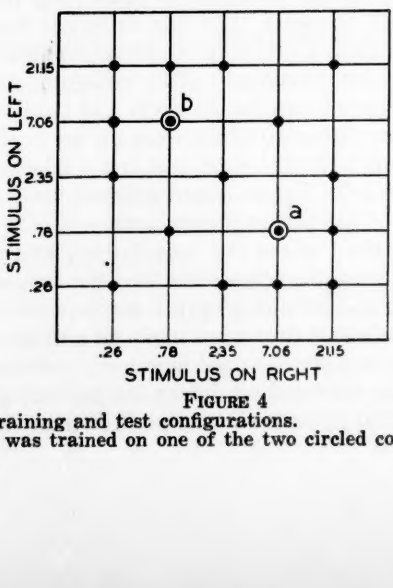


FIGURE 4

Plot of the training and test configurations.

Each animal was trained on one of the two circled configurations and tested on the other 16.

dow and a black card in the other. The positions of the two cards were alternated in random fashion. Not until the animal mastered the problem of jumping to the white card was formal training with the stimuli to be discriminated begun. But training the animals to jump to a white card when it is paired with a black one has the effect of training them to jump to the brighter of the two cards. This procedure, then, gives training on a light-dark discrimination before any formal records of the experiment are taken.

Because some such training seems necessary with the Lashley jumping procedure, we used it with half of our animals. With the other half, we reversed the preliminary training and made the animals jump to the black card and avoid the white one. The reversal of the usual technique had, of course, the effect of training these animals to choose the darker in preference to the lighter stimulus during preliminary training.

Twelve animals were given each type of preliminary training. Half of each of these two groups was presented on all regular trials with configuration *a* of figure 3 and half with configuration *b* of figure 3. Half of each of these four groups was trained to jump to the right and half to jump to the left. There were then eight different combinations of preliminary training, configuration, and response. Three animals were trained on each combination. These eight groups, with the preliminary and training conditions summarized for each are described in the first seven columns of table 1.

Training was continued at the rate of 10 trials per day until the animal had made no errors on five consecutive days. If an animal made errors on the first trial of the first day, but made none thereafter, the first 5 days were accepted as satisfying the learning criterion. Learning was so rapid that this criterion was met by most animals in five or six days (including the criterion runs).

Transfer tests were introduced after satisfying the learning criterion. These tests were given on the sixth and twelfth trials of each day's series. The remaining 10 trials were on the training configuration. The animals were rewarded on test trials whether they jumped to the right or to the left. The different animals were given from two to six tests on each of the test configurations.

Column 9 of table 1 gives the frequencies, and percentages, of jumps in agreement with the hypothesis that the responses were made on a relational basis and column 10 gives corresponding data with respect to the hypothesis that they were made on a directional basis. The basis which each rat apparently used is named in column 11. A directional hypothesis was recorded whenever the percentage of test jumps in agreement with that hypothesis was greater than the percentage of

1	2	3	4	5	6	7	8	9	10	11
Group No.	An. No.	Preliminary training positive to	Training configuration see Figure 3	Response required to configuration	Runs to completion of learning criterion	Errors to completion of learning criterion*	Number of test trials†	No. & % right on relative hypothesis‡	No. & % right on directional hypothesis‡	Apparent basis of response on tests§
1	1	Light	a	R	50	0	48(33)	28(85)	18(37)	Rel.
	2	Light		R	50	1(1)	48(33)	33(100)	30(63)	Rel.
	3	Light		R	50	2(1)	48(33)	28(85)	24(50)	Rel.
2	4	Light	b	L	50	0	48(33)	20(61)	42(88)	Dir.
	5	Light		L	50	0	48(33)	30(91)	32(67)	Rel.
	6	Light		L	70	2(2)	48(33)	30(91)	19(40)	Rel.
3	7	Light	b	R	50	7(1)	34(23)	11(48)	34(100)	Dir.
	8	Light		R	60	6(2)	34(23)	11(48)	34(100)	Dir.
	9	Light		R	60	18(3)	34(23)	11(48)	34(100)	Dir.
4	10	Light	a	L	50	5(1)	34(25)	13(52)	34(100)	Dir.
	11	Light		L	60	3(2)	34(24)	12(50)	34(100)	Dir.
	12	Light		L	60	4(2)	34(24)	12(50)	34(100)	Dir.
5	13	Dark	a	R	60	4(2)	48(33)	15(45)	48(100)	Dir.
	14	Dark		R	60	2(1)	48(33)	15(45)	48(100)	Dir.
	15	Dark		R	60	2(2)	48(33)	15(45)	48(100)	Dir.
6	16	Dark	b	L	60	2(1)	48(33)	15(45)	48(100)	Dir.
	17	Dark		L	60	2(2)	62(47)	26(55)	59(95)	Dir.
	18	Dark		L	50	1	48(33)	15(45)	48(100)	Dir.
7	19	Dark	b	R	50	0	66(51)	21(41)	60(91)	Dir.
	20	Dark		R	60	1	67(52)	18(35)	64(96)	Dir.
	21	Dark		R	60	2(2)	96(66)	57(86)	44(46)	Rel.
8	22	Dark	a	L	50	1	48(33)	15(45)	48(100)	Dir.
	23	Dark		L	50	1	48(33)	15(45)	48(100)	Dir.
	24	Dark		L	60	1	96(66)	60(91)	51(53)	Rel.

* The first number gives the number of individual wrong responses during training; the number in parentheses gives the number of different trials on which wrong responses were made.

† The first number is the total number of test trials. The number in parentheses is the number of test trials using configurations made up of different stimuli, e.g., circles 5 and 1 or 2 and 3. The difference between these two numbers is the number of test trials using configurations made up of two circles of the same size, e.g., 1 and 1 or 4 and 4.

‡ The first number gives the total number of test responses in agreement with the hypothesis stated at the top of the column; the number in parentheses gives the percentage of all test responses which were in agreement with the hypothesis stated at the top. In computing the percentage of test responses in agreement with the relational hypothesis, the number of tests composed of different stimuli was used as a base.

§ A directional hypothesis is recorded when the percentage in column 10 is higher than the corresponding percentage in column 9. When the reverse is true a relational hypothesis is recorded.

test jumps in agreement with the relational hypothesis. Whenever the reverse was true, a relational hypothesis was recorded.

When the preliminary training required the animal to jump to the white card and the regular training required it to jump to the side of the larger circle (Groups 1 and 2), or, when the preliminary training required the animal to jump to the black card and the regular training required it to jump to the side of the smaller circle (Groups 7 and 8), the preliminary and regular training both reinforced a tendency to respond to the larger (brighter) or smaller (dimmer) of the two stimuli. Under these conditions, seven of the 12 animals of groups 1, 2, 7, and 8 responded to the test configurations on a relational basis. The other five responded on a directional basis.

When the preliminary training required the animal to jump to the white card and the regular training required it to jump to the side of the smaller circle (Groups 3 and 4), or, when the preliminary training required the animal to jump to the black card and the regular training required it to jump to the side of the larger circle (Groups 5 and 6), the regular training conflicted with any tendency, which may have been established by the preliminary training, to respond on a relational basis. Under these conditions all 12 animals of Groups 3, 4, 5, and 6 responded to the test configurations on a directional basis. (Of the 506 test jumps, 503 were to the side which had been correct during regular training.)

The difference in the behavior of groups trained to go to the white square and those trained to go to the black square indicate that this preliminary training, which is normally not recorded as part of the experimental data, is important in determining how the rat will respond on the subsequent regular training. This influence has hitherto been neglected in most studies of discrimination.

The results as a whole indicate that a directional habit was the more frequent type of response learned under these experimental conditions. Seventeen of the 24 animals responded to the test configurations on a directional basis and only seven on a brightness basis. In conclusion, it seems considerably safer to assume that the rats' initial response in the discrimination problem is directional rather than relative. Consequently, *we will define the response directionally as a jump to the right or a jump to the left.*

The Two-Configuration Problem

On the basis of the definitions of the stimulus and of the response given in the preceding section, an equation of the learning curve can be derived. We will consider first the case in which the problem is to discriminate between two configurations each of which is composed of

two stimuli differing with respect to one variable. The animal may be rewarded for jumping to the left in response to configuration *a* and to the right in response to configuration *b* of figure 1. This is the situation which would usually be described as requiring the animal to respond positively to the larger of the two stimuli.

Definitions of Variables

As far as possible we have used the symbols previously employed by Thurstone (28, 29) and Gulliksen (4) in order to emphasize the relationship between the concepts common to the three equations. The symbols e, s, u, w, h, g, k and c have been taken from the previous equations. We have added subscripts and primes to indicate extensions of meaning necessary in the present more generalized formulation.

If we designate a cumulative count of correct responses by w and a cumulative count of incorrect responses by u, and assume that the subject is being trained to go to the *left* for configuration a, and to the *right* for configuration b (see Figure 1), we have:

w_a = number of *left* or *correct* responses to configuration a. The animal is *rewarded* for each of these responses.

u_a = number of *right* or *incorrect* responses to configuration a. The animal is *punished* for each of these responses.

Correspondingly,

w_b = number of *right* or *correct* responses to configuration b. The animal is *rewarded* for each of these responses.

u_b = number of *left* or *incorrect* responses to configuration b. The animal is *punished* for each of these responses.

The strengths of the tendencies to make these four responses may be represented as follows:

s_a = strength of the tendency to make a *left* or *correct* response to configuration a.

e_a = strength of the tendency to make a *right* or *incorrect* response to configuration a.

s_b = strength of the tendency to make a *right* or *correct* response to configuration b.

e_b = strength of the tendency to make a *left* or *incorrest* response to configuration b.

If the animal is trained positively to the smaller of the two stimuli (See configuration a and b in figure 1), the words "right" and "left" would be interchanged in all of the foregoing definitions. Such a change would not affect any of the subsequent developments.

a b

CR = left CR = right

Basic Assumptions

(1) It will be assumed that the strength of the tendency to go to the *right* is influenced by the amount of punishment and reward that the rat has received for making *right* choices, *i.e.,*

right

$$e_a = f(u_a, w_b) \text{ and}$$

$$s_b = f(w_b, u_a).$$

Correspondingly, the strength of the tendency to go to the *left* is a function of the punishment and reward of *left* choices, *i.e.,*

$$s_a = f(w_a, u_b) \text{ and}$$

$$e_b = f(u_b, w_a).$$

More definitely it will be assumed that:

$$\frac{\delta s_a}{\delta w_a} = k \quad \text{and} \quad \frac{\delta s_b}{\delta w_b} = k \cdot$$

k = the amount by which each *reward* of a *correct* response *increases* the strength of the tendency to make the same response to the *same* configuration.

$$\frac{\delta s_a}{\delta u_b} = -c' \quad \text{and} \quad \frac{\delta s_b}{\delta u_a} = -c' \cdot$$

c' = the amount by which each *punishment* of an *incorrect* response to one configuration *decreases* the tendency to make the *same* (or *correct*) response to the other configuration ($c' < c$).

$$\frac{\delta e_a}{\delta u_a} = -c \quad \text{and} \quad \frac{\delta e_b}{\delta u_b} = -c \cdot$$

c = the amount by which each *punishment* of an *incorrect* response *decreases* the strength of the tendency to make the same response to the *same* configuration.

$$\frac{\delta e_a}{\delta w_b} = k' \quad \text{and} \quad \frac{\delta e_b}{\delta w_a} = k' \cdot$$

k' = the amount by which each *reward* of a *correct* response to one configuration *increases* the tendency to make the *same* (or *incorrect*) response to the *other* configuration ($k' < k$).

The parameters k and c represent the learning constants of an animal — the amount by which it profits from each rewarded or punished trial. The parameters k' and c' represent those learning constants diminished by the effect of the distance separating the two con-

figurations. k' and c' are, then, respectively, functions of k and distance and c and distance. The relation between the size of k' or c' and the distance between the two configurations may be thought of as a generalization gradient (9, 13). This concept is adapted from Pavlov's concept of generalization. The greater the distance between the two configurations, the less is the value of k' or c'. The nature of the function relating amount of generalization to distance has not been determined for visual discrimination data. Hovland (9) has found a negatively accelerated generalization gradient for a galvanic response conditioned to tonal stimuli. But the closest test stimulus used by him was 25 j. n. d.'s removed from his training stimulus. The shape of the generalization gradient between the training stimulus and one removed by 25 j. n. d.'s has not been determined. Perhaps the entire gradient might be represented by $k' = \frac{k}{kd^2 + 1}$. For the development of the present theory it is not necessary to determine the exact function relating k' (or c') to distance. It need only be assumed that the function is a monotonic decreasing one. The findings of both Bass and Hull (1) and Hovland (9) support this assumption. When stimulus configurations are accurately scaled and quantitative predictions of the difficulty of different problems are attempted, it will become necessary to determine the shape of the generalization gradient more precisely.

Solving each pair of partial differential equations involving the same response tendency (*i.e.*, the same dependent variable) gives the following equations for the four functions indicated above:

$$e_a = g_a - cu_a + k'w_b\,, \qquad \text{(II-1)*}$$

$$s_a = h_a - c'u_b + kw_a\,, \qquad \text{(II-2)}$$

$$e_b = g_b - cu_b + k'w_a\,, \qquad \text{(II-3)}$$

$$s_b = h_b - c'u_a + kw_b\,. \qquad \text{(II-4)}$$

The g's and h's are constants of integration and represent the strength of each of the tendencies at the beginning of the experiment.

(2) It will be assumed that the strengths of the correct and incorrect responses are related to the number of these responses by the equations:

$$\frac{du_a}{dw_a} = \frac{e_a}{s_a}\,, \qquad \text{(II-5)}$$

$$\frac{du_b}{dw_b} = \frac{e_b}{s_b}\,. \qquad \text{(II-6)}$$

* The number of each equation dealing with the two-configuration case will be preceded by II.

Substituting equations (II-1), (II-2), (II-3), and (II-4) in (II-5) and (II-6) gives the following differential equations:

$$\frac{du_a}{dw_a} = \frac{g_a - cu_a + k'w_b}{h_a - c'u_b + kw_a}, \tag{II-7}$$

$$\frac{du_b}{dw_b} = \frac{g_b - cu_b + k'w_a}{h_b - c'u_a + kw_b}. \tag{II-8}$$

(3) A third relationship is found in the experimental conditions. These are usually so arranged that either:

a) each of the configurations a and b is presented the same number of times; or
b) the animal makes the same number of correct responses to each configuration.

The former is represented by

$$u_a + w_a = u_b + w_b\,, \tag{II-9}$$

the latter by

$$w_a = w_b\,. \tag{II-10)*}$$

Lashley

Equation (II-9) represents the experimental conditions under which each configuration is presented an equal number of times. Only one response is allowed to each configuration. This technique is used by a number of experimenters. Equation (II-10) represents the experimental conditions under which the animal is allowed to respond to each configuration until it makes the correct response. The total number of correct responses is thus the same to both configurations, but the number of wrong responses to one may not equal the number of wrong responses to the other. This technique is sometimes known as the Lashley method (19).

The system of equations represented by either

(II-7), (II-8), and (II-9)

or by

(II-7), (II-8), and (II-10)

* It is possible to regard the u's and w's as functions of time. In this case (II-10) is exactly given by

$$w_a(t_i) = w_b(t_i) \tag{II-10a}$$

for a large number of values of i throughout the course of the experiment. With the qualification that $| w_a(t) - w_b(t) | \leqslant \varepsilon$ for every t. Where ε is small (as is always the case with the Lashley technique) we can approximate (11-10a) by (II-10).

Since w_a and w_b are separate quantities in the learning process of the animal, replacing them by a single variable is not only an approximation to the solutions of (II-7) and (II-8) but also an approximation of the psychological model represented by (II-7) and (II-8).

A similar argument applies in the case of equation (II-9).

is a soluble set of equations, but there is no closed solution and at present a series approximation is not worth while. Therefore, one simplifying assumption will be made. We will consider only those cases where the initial strengths of the two tendencies to respond correctly equal each other

$$h_a = h_b \tag{II-11}$$

and the initial strengths of the two tendencies to respond incorrectly equal each other

$$g_a = g_b\ ; \tag{II-12}$$

h_a need not equal g_a.

With this limitation it can be seen that equation (II-8) may be obtained from equation (II-7) by interchanging the subscripts a and b. It can be shown that with this limitation and with equation (II-10) it follows that

$$u_a = u_b\,. \tag{II-13}$$

This can be seen by expanding u_a and u_b as power series (20) in the variable w:

$$u_a = \sum_{i=0}^{\infty} a_i w^i\,, \qquad u_b = \sum_{i=0}^{\infty} b_i w^i\,. \tag{II-14}$$

Substituting (II-14) in (II-7) and (II-8), assuming (II-11) and (II-12), and comparing coefficients of like powers of w, we find that

$$a_i = b_i$$

for every value of i, which proves (II-13).

Using (II-10) and (II-13), equations (II-7) and (II-8) reduce to one equation with the subscripts eliminated, leaving:

$$\frac{du}{dw} = \frac{g - cu + k'w}{h - c'u + kw}\,. \tag{II-15}$$

strength of correct + incorrect responses are relative not [illegible] [illegible]

Derivation of Equation of the Learning Curve

Equation (II-15) may be solved by first shifting the origin to remove the constants g and h. In order to do this, set

$$u = u' + m \tag{II-16}$$

and

$$w = w' + n \tag{II-17}$$

where

$$m = \frac{kg - k'h}{kc - k'c'}\,, \tag{II-18}$$

$= \frac{(kg - k'h)}{k^2 - k'^2}$

$(k + k')(k - k')$

and

$$n = \frac{c'g - ch}{kc - k'c'}. \tag{II-19}$$

Differentiating (II-16) and (II-17) gives

$$\frac{du}{dw} = \frac{du'}{dw'}. \tag{II-20}$$

Substituting (II-16), (II-17), (II-18), (II-19), and (II-20) in (II-15) gives:

$$\frac{du'}{dw'} = \frac{g - c[u' + \frac{kg - k'h}{kc - k'c'}] + k'\,[w' + \frac{c'g - ch}{kc - k'c'}]}{h - c'[u' + \frac{kg - k'h}{kc - k'c'}] + k\,[\,w' + \frac{c'g - ch}{kc - k'c'}]}. \tag{II-21}$$

Simplifying (II-21), gives

$$\frac{du'}{dw'} = \frac{k'w' - cu'}{kw' - c'u'}. \tag{II-22}$$

Again defining a new variable v by the equation

$$u' = vw' \tag{II-23}$$

and differentiating (II-23) we have

$$\frac{du'}{dw'} = v + w'\,\frac{dv}{dw'}. \tag{II-24}$$

Substituting (II-23) and (II-24) in (II-22) and cancelling w' from the right-hand member of the equation gives

$$v + w'\,\frac{dv}{dw'} = \frac{k' - cv}{k - c'v}. \tag{II-25}$$

Separating the variables in (II-25) gives

$$\frac{dw'}{w'} = \frac{k - c'v}{k' - (k + c)\,v + c'v^2}\,dv. \tag{II-26}$$

To integrate the right-hand member of (II-26) it is necessary to multiply numerator and denominator by 2 and to add and subtract $c\,dv$ in the numerator. Then, rearranging terms, we may write

$$\frac{dw'}{w'} = (1/2)\,\frac{k + c - 2c'v}{k' - (k+c)\,v + c'v^2}\,dv$$

$$+\left(\frac{k-c}{2}\frac{1}{k'-(k+c)v+c'v^2}\right)dv\,. \tag{II-27}$$

integrating (II-27) gives

$$\log w' = -1/2\log\left[c'v^2-(k+c)v+k'\right]+\left(\frac{k-c}{2}\right)\left(\frac{1}{r}\right)$$

$$\log\left[\frac{2c'v-(k+c)-r}{2c'v-(k+c)+r}\right]+\log K \tag{II-28}$$

where

$$r^2=(k+c)^2-4k'c'\,, \tag{II-29}$$

and log K is the constant of integration. The logarithmic rather than the arc tangent form of the integral for the last term of equation (II-27) is appropriate since $r^2 > 0$.

Taking antilogarithms of (II-28) gives

$$w' = K\left\{\frac{1}{[c'v^2-(k+c)v+k']^{\frac{1}{2}}}\right\}\left\{\frac{2c'v-(k+c)-r}{2c'v-(k+c)+r}\right\}^{\frac{k-c}{2r}}. \tag{II-30}$$

Substituting the original variables u and w from equations (II-16), (II-17), and (II-23) gives

$$w-n = K\left[c'\left(\frac{u-m}{w-n}\right)^2-(k+c)\frac{u-m}{w-n}+k'\right]^{-\frac{1}{2}}$$

$$\times\left[\frac{2c'\dfrac{u-m}{w-n}-(k+c)-r}{2c'\dfrac{u-m}{w-n}-(k+c)+r}\right]^{\frac{k-c}{2r}} \tag{II-31}$$

where m, n and r are defined by equations (II-18), (II-19), and (II-29).

Equation (II-31) may be rewritten as follows:

$$\left(\frac{w-n}{K}\right)^2 = \left[\frac{(w-n)^2}{c'(u-m)^2-(k+c)(u-m)(w-n)+k'(w-n)^2}\right]$$

$$\times\left[\frac{2c'(u-m)-(k+c+r)(w-n)}{2c'(u-m)-(k+c-r)(w-n)}\right]^{\frac{k-c}{r}} \tag{II-32}$$

or

$$[c'(u-m)^2-(k+c)(u-m)(w-n)+k'(w-n)^2]$$

$$\times\left[\frac{2c'(u-m)-(k+c-r)(w-n)}{2c'(u-m)-(k+c+r)(w-n)}\right]^{\frac{k-c}{r}} = K^2\,. \tag{II-33}$$

Since at the origin $u = w = 0$, we may write K^2 from equation (II-33) by setting u and w both equal to zero. Doing this gives

$$K^2 = \left[c'm^2 - (k+c)mn + k'n^2\right]\left[\frac{n(k+c-r) - 2c'm}{n(k+c+r) - 2c'm}\right]^{\frac{k-c}{r}} . \tag{II-34}$$

Equation (II-33), giving K^2 the value defined by (II-34), is the solution for the conditions given at the beginning of the derivation.

If c' and k' in equation (II-33) are set equal to zero (i.e., if transfer effects are negligible) the equation simplifies to the form previously presented by Gulliksen (4). If, further, $c = k$ in equation (II-33) (i.e., if the effect of punishing an incorrect response is assumed to be equal to the effect of rewarding a correct response) the equation further reduces to the form developed by Thurstone (29). Both of these equations are, therefore, special cases of equation (II-33).

Equation (II-33) may be simplified in a special case by ignoring the exponential term. As c approaches k this term approaches unity, so that (II-33) reduces to

$$c'(u-m)^2 - (k+c)(u-m)(w-n) + k'(w-n)^2 = K^2. \tag{II-35}$$

Re-evaluating K^2 by setting $u = w = 0$ gives

$$K^2 = c'm^2 - (k+c)mn + k'n^2 . \tag{II-36}$$

Substituting (II-36) in (II-35) and simplifying gives

$$c'u^2 - 2c'mu - (k+c)(uw - mw - nu) + k'w^2 - 2k'nw = 0 . \tag{II-37}$$

Substituting the values of m and n from (II-18) and (II-19) gives

$$c'u^2 + k'w^2 - (k+c)uw + \left[\frac{(k+c)(c'g-ch)}{kc-k'c'} - \frac{2c'(kg-k'h)}{kc-k'c'}\right]u$$
$$+\left[\frac{(k+c)(kg-k'h)}{kc-k'c'} - \frac{2k'(c'g-ch)}{kc-k'c'}\right]w = 0 . \tag{II-38}$$

Equation (II-38) holds exactly in the special case where $c = k$ and $c' = k'$ since then the exponential term in equation (II-33) is equal to unity. In this case equation (II-38) becomes

$$u^2 + w^2 - 2\frac{k}{k'}uw - 2\frac{h}{k'}u + 2\frac{gw}{k'} = 0 . \tag{II-39}$$

Assuming in addition that the strengths of the two responses involved are equal at the beginning, *i.e.*, that $g = h$, we have

$$u^2 + w^2 - 2\frac{k}{k'}uw - 2\frac{g}{k'}u + 2\frac{g}{k'}w = 0 . \tag{II-40}$$

Solving explicitly for u we get

$$u = \frac{g}{k'} + \frac{k}{k'}w - \sqrt{\left[\left(\frac{k}{k'}\right)^2 - 1 \right] w^2 + \left[\frac{k}{k'} - 1 \right] \frac{2g}{k'} w + \left[\frac{g}{k'} \right]^2} \; . \tag{II-41}$$

Equations (II-38), (II-39), (II-40), and (II-41) are hyperbolas passing through the origin. The curve is shown in figure 5. The terms in these equations have the following meanings.

u = cumulative count of incorrect responses.
w = cumulative count of correct responses.
g = initial strength of the tendency to respond incorrectly.
h = initial strength of the tendency to respond correctly.
k = the amount by which each reward of a response increases the strength of the tendency to make the same response to the same configuration.
k' = the amount by which each reward of a response increases the strength of the tendency to make the same response to the other configuration.
c = the amount by which each punishment of a response decreases the strength of the tendency to make the same response to the same configuration.
c' = the amount by which each punishment of a response decreases the strength of the tendency to make the same response to the other configuration.

Properties of This Learning Curve

In equation (II-38) the two asymptotes are

$$u = \frac{k + c \pm r}{2c'} w + \left[m - \frac{(k + c \pm r)n}{2c'} \right]. \tag{II-42}$$

The secant of the angle between the two asymptotes is

$$\frac{\sqrt{(k+c)^2 + (c'-k')^2}}{c' + k'}.$$

The tangent of this angle is $\frac{r}{c' + k'}$. $r^2 = (k+c)^2 - 4k'c'$

The asymptotes intersect at the point $u = m$, $w = n$.

In the special case $g = h$, $c = k$, and $c' = k'$ for which equation (II-40) holds exactly, the asymptotes are

$$u = \left[\frac{k \pm \sqrt{k^2 - k'^2}}{k'}\right] w + \frac{g}{k'}\left[1 \pm \sqrt{\frac{k - k'}{k + k'}}\right]. \qquad \text{(II-43)}$$

The secant of the angle between the asymptotes is $\dfrac{k}{k'}$.

The tangent of this angle is $\dfrac{\sqrt{k^2 - k'^2}}{k'}$.

The asymptotes intersect at the point $u = \dfrac{g}{k + k'}$, $w = -\dfrac{g}{k + k'}$.

Sec ∡A = C/AD

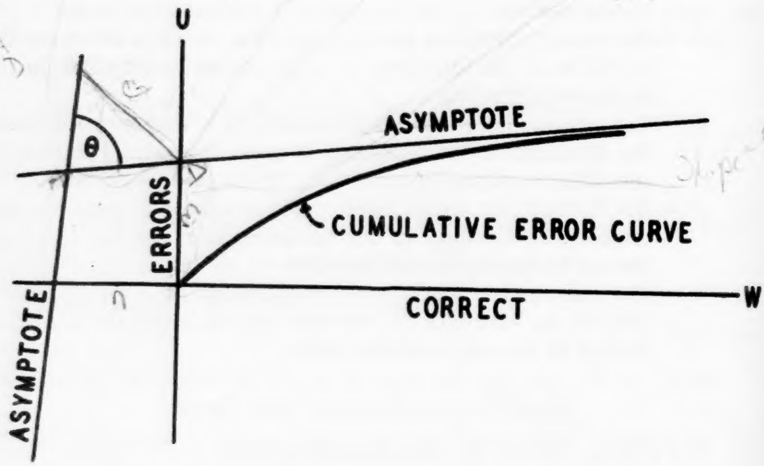


FIGURE 5
The cumulative error curve and its asymptotes.

Application to Experimental Data

(1) In deriving equation (II-39) from (II-38) it was assumed that $c = k$ and $c' = k'$. For any given set of data it is possible to test this assumption.

If $c = k$ and if $c' = k'$, then the coefficients of u^2 and w^2 in equation (II-37) will be equal. After these coefficients have been found for a particular learning curve, their equality or departure from equality will reveal the equality or inequality of the learning effects of punishing an incorrect and rewarding a correct response.

(2) In deriving equation (II-40) from equation (II-39), it was

assumed that the strengths of the two responses were equal at the beginning of practice. It is possible to check upon the accuracy of this assumption for any given set of data, after it has been shown that $c' = k'$.

The asymptotes of the curve (II-39), intersect at the point $\frac{k'g - kh}{(k^2 - k'^2)}$ and $\frac{kg - k'h}{(k^2 - k'^2)}$. If g and h are in reality equal, these coordinates reduce to $\frac{g}{k + k'}$ and $\frac{-g}{k + k'}$. If the intersection of the asymptotes of the plotted curve is equidistant from the u and w axes, then g and h must be equal to each other. Any departure from equality will indicate which tendency was stronger. The amount of inequality will indicate the size of the difference in the initial strengths of the two tendencies.

(3) It may be deduced that: *The difficulty of a problem, as measured by the maximum level of accuracy attainable, is inversely related to the distance separating the two configurations to be discriminated.*

From equation (II-42) or equation (II-43) — the equations of the asymptotes of the learning curve — it can be seen that the slope of the asymptotes is a function of the relative sizes of k and k' and of c and c'. These ratios, in turn, are functions of the distance separating the configurations to be discriminated. If the distance is zero, a situation which can occur only when the subject is forced to attempt to discriminate between a configuration and itself, then $c = c'$ and $k = k'$. In this case the two asymptotes will have a slope of unity, the learning curve equation will become $u = w$, no learning will be possible, and the subject will continue to respond with only chance accuracy.

From the same equations it can be seen that if the distance separating the two configurations is sufficiently great to make c' and k' insignificantly different from zero, then the asymptote approached by the learning curve will have a slope of zero. In this case perfect learning is possible.

For most configurations used in ordinary learning experiments c' and k' have probably been small in comparison with c and k. No learning studies have reported attempts to determine whether or not the asymptotes had a small positive slope which increased with the difficulty of the discrimination, but much of the psychophysical work has shown that the level of accuracy attainable increases with an increase in the difference between the stimuli being discriminated.

REFERENCES

1. BASS, M. J., and HULL, C. L., "The irradiation of a tactile conditioned reflex in man," *Jour. Comp. Psychol.*, 1934, **17**, 47-65.
2. FIELDS, P. E., "Studies in concept formation: II. A new multiple stimulus jumping apparatus for visual figure discrimination," *Jour. Comp. Psychol.*, 1935, **20**, 183-203.
3. GULLIKSEN, H., "Studies of transfer of response: I. Relative versus absolute factors in the discrimination of size by the white rat," *Ped. Sem. and Jour. Genet Psychol.*, 1932, **40**, 37-51.
4. GULLIKSEN, H., "A rational equation of the learning curve based on Thorndike's Law of Effect," *Jour. Gen. Psychol.*, 1934, **11**, 395-434.
5. GULLIKSEN, H., "The relationship between degree of original learning and degree of transfer," *Psychometrika*, 1936, **1**, 37-43.
6. HEBB, D. O., "The innate organization of visual activity: II. Transfer of response in the discrimination of brightness and size by rats reared in total darkness," *Jour. Comp. Psychol.*, 1937, **24**, 277-299.
7. HEBB, D. O., "Innate organization of visual activity: III. Discrimination of brightness after removal of the striate cortex in the rat," *Jour. Comp. Psychol.*, 1938, **25**, 427-437.
8. HELSEN, H., "Insight in the white rat," *Jour. Exper. Psychol.*, 1937, **10**, 378-396.
9. HOVLAND, C. I, "The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone," *Jour. Gen. Psychol.*, 1937, **17**, 125-148.
10. HOVLAND, C. I., "The generalization of conditioned responses: IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses," *Jour. Exper. Psychol.*, 1937, **21**, 261-276.
11. HULL, C. L., "The conflicting psychologies of learning—A way out," *Psychol. Rev.*, 1935, **42**, 491-516.
12. HULL, C. L., "Mind, mechanism, and adaptive behavior," *Psychol. Rev.*, 1937, **44**, 1-32.
13. H(ULL), C. L., "Some tentative theorems concerning stimulus equivalence. (Memorandum to Mr. Hull's graduate seminar)," Mimeographed, Feb. 14, 1938, pp. 10.
14. KÖHLER, W., *Gestalt Psychology*. New York: Liveright, 1929, $x + 403$.
15. KRECHEVSKY, I., " 'Hypotheses' vs. 'chance' in the pre-solution period in sensory discrimination-learning," *Univ. Calif. Publ. Psychol.*, 1932, **6**: No. 3, 27-44
16. KRECHEVSKY, I., "A study of the continuity of the problem-solving process," *Psychol. Rev.*, 1938 **45**, 107-133.
17. LASHLEY, K. S., *Brain mechanisms and intelligence*, Chicago: University of Chicago Press, 1929, *xiv* + 186 + 11 plates.
18. LASHLEY, K. S., "The mechanism of vision: I. A method for rapid analysis of pattern-vision in the rat," *Jour. Genet. Psychol.*, 1930, **37**, 453-460.
19. LASHLEY, K. S., "The mechanism of vision: XV. Preliminary studies of the rat's capacity for detail vision," *Jour. Gen. Psychol.*, 1938, **18**, 123-193.
20. MOULTON, F. R., *Differential Equations*. New York: Macmillan, 1930, *xv* + 395.
21. MUNN, N. L., "An apparatus for testing visual discrimination in animals," *Jour. Genet. Psychol.*, 1931, **39**, 342-358.

22. PERKINS, F. T., and WHEELER, R. H., "Configurational learning in the goldfish," *Comp. Psychol. Monog.*, 1930, **7**: No. 31, pp.50.
23. RASHEVSKY, N., *Mathematical Biophysics*, Chicago: University of Chicago Press, 1938, *xviii* + 340.
24. RÉVÉSZ, G., "Experimental study in abstraction in monkeys," *Jour. Comp. Psychol.*, 1925, **5**, 293-343.
25. SPENCE, K. W., "The nature of discrimination learning in animals," *Psychol. Rev.*, 1936, **43**, 427-449.
26. SPENCE, K. W., "The differential response in animals to stimuli varying within a single dimension," *Psychol. Rev.*, 1937, **44**, 430-444.
27. THORNDIKE, E. L., *The fundamentals of learning*, New York: Teacher's College, Columbia Univ., Bur. Publ., 1932, *xvii* + 638.
28. THURSTONE, L. L., "The learning function," *Jour. Gen. Psychol.*, 1930, **3**, 469-493.
29. THURSTONE, L. L., "The error function in maze learning," *Jour. Gen. Psychol.*, 1933, **9**, 288-301.
30. WARDEN, C. J., and ROWLEY, J. B., "The discrimination of absolute versus relative brightness in the ring dove, *Turtur risorius*," *Jour. Comp. Psychol.*, 1929, **9**, 317-337.
31. WARDEN, C. J., and WINSLOW, C. N., "The discrimination of absolute versus relative size in the ring dove, *Turtur risorius*," *Jour. Genet. Psychol.*, 1931, **39**, 328-341.
32. WILEY, L. E., and WILEY, A. M., "Studies in the learning function," *Psychometrika*, 1937, **2**, 1-19, 107-120, 161-164.
33. WOLFLE, D. L., "Absolute brightness discrimination in the white rat," *Jour. Comp. Psychol.*, 1937, **24**, 59-71.
34. YERKES, R. M., and WATSON, J. B., "Methods of studying vision in animals," *Behavior Monographs*, 1911, **1**: No. 2, pp. iv + 90.
35. YOUNG, G., and HOUSEHOLDER, A. S., "Discussion of a set of points in terms of their mutual distances," *Psychometrika*, 1938, **3**, 19-22.

PSYCHOMETRIKA—VOL. 3, NO. 3
SEPTEMBER, 1938

THE UNIT HIERARCHY AND ITS PROPERTIES

CYRIL BURT
Psychological Laboratory, University College, London

A correlation matrix may be expanded as the weighted sum of a series of 'unit hierarchies'. The properties of the 'unit hierarchy' are not only of theoretical interest for themselves, but lead to simpler modes of practical calculation. The analysis is analogous to a spectral set of projective operations in quantum-theory: and the analogy itself suggests many further problems and solutions.

In several investigations (1, 2, 3, 21) my research students and I have applied a modified form of factor analysis that has proved to possess definite advantages of its own. It seems a little more exact in theory than the methods generally current in England, a little less laborious in practice than those recently advocated in America, and rather more in line with solutions worked out by mathematicians and physicists who deal with similar issues in other fields. Hitherto, however, the references to it have been incidental to some special problem. Accordingly, I gladly welcome the suggestion that a more explicit account should be offered of the theoretical arguments on which it rests.

I. Definition of Factors

In summarizing the general principles available for analyzing a table of correlations, I have distinguished two main alternatives: (a) a 'group factor' method, and (b) a 'common factor' method. Method (a) proceeds by partitioning the matrix to be factorized into separate submatrices, and seeks to explain each of these by a positive factor of limited range; method (b) deals with the matrix throughout as a whole, even when some of its elements or submatrices are negative. The latter mode of approach has usually been preferred by the English School, since its members have been interested primarily in the search for a single central factor. But in early work on educational testing, I found it necessary from the very outset to envisage, not only 'general factors', but also 'specific' — i.e., factors apparently restricted to groups of tests; accordingly, the 'group factor' method was introduced for purposes of practical calculation. Since in theory the group factors (which need not be independent) may always be obtained from the independent common factors by a further rotation

of axes, I shall confine myself here to what I have called method (b).

This mode of reduction implies that the factors so obtained can claim in the first instance to be nothing more than mathematical abstractions; as they stand, they will have no necessary relation to concrete psychological distinctions. If in actual practice such identifications are often possible, that is due to an appropriate choice of the tests to be correlated, not to any properties of the mathematical procedure as such. What the mathematical procedure can guarantee is that each factor, as it is extracted, will have a maximum discriminative and predictive value, but not that it will have a maximum of psychological meaning or of psychological simplicity.*

In accordance with these principles I define the k^{th} factor as that component which will account for the maximum amount of variance remaining after $(k - 1)$ components have been removed. This definition implies that our primary aim is the 'analysis of variance'. Thus I consider the real object of factor-analysis to be the analysis of the original matrix of measurements; the analysis of the correlations (which originally gave rise to the search for factors) I regard as merely a means to that end. Such a standpoint is in keeping with the trend of English statistical work in other directions (cf. Fisher, 4); and enables us to apply factor-analysis to the results of correlating persons as well as traits.†

* As a working procedure, the group factor method is required only when the variables correlated form a discontinuous series, that is to say, when the coefficients can be grouped together to form two or more submatrices of comparatively high positive figures separated by submatrices of negative, zero, or comparatively low figures. A two-fold pattern of this nature occurs most conspicuously when the first general factor has been partialled out; but it may also be seen in the observed correlations themselves, when the general factor has in effect been eliminated by some other device (e.g., by choosing homogeneous populations or material, or by re-scaling the measurements transversely in terms of deviations about their average). Even then the whole table may still be analyzed in terms of a common factor, which will now be bipolar. (One or two recent investigators have assumed that it is impossible to extract a general factor when one or more rectangular submatrices contain negative coefficients: the method described below shows clearly that this assumption is erroneous). In certain circumstances, however, the variables selected for correlation form discontinuous groups: for instance, in correlating the results of scholastic tests, the tests of (i) handwork, (ii) arithmetic, (iii) verbal subjects (reading, spelling, composition), yield three discontinuous blocks of positive correlations. In such a case the group factor method is more appropriate. A typical example may be seen in my analysis of school subjects in *Distribution and Relations of Educational Abilities*, 1917, Table XX, p. 57.

† These points were briefly stated in the memorandum I was asked to prepare for the International Institute Examinations Inquiry (1, pp. 247, 275, 306). Since that was first drawn up, Kelley (5) has published a 'new method of analysis' which is also to be applied to covariances: but he seems to be in error in assuming that the results are identical whether covariances or correlations are used. In what follows, since the more appropriate terms, 'comultiplying' and 'covariating', are still somewhat unfamiliar, and the distinctions are here unimportant, I shall use the term 'correlating' loosely to cover all three processes.

II. The Canonical Expansion

With the foregoing definition of factors, the problem of factor analysis becomes simply a problem in multiple correlation; and a unique algebraic solution can be at once obtained along the usual lines by invoking the principle of least squares.* If there are n tests or traits, the result will express the matrix of observed measurements in terms of n independent factors, which emerge in order of the amount contributed by each to the total variance. Of these, however, the factors that emerge last of all will have little or no statistical significance. If, therefore, we seek only the first k, and from these reconstruct a reduced matrix of test-measurements, or a reduced matrix of correlations, then the reconstruction will yield the nearest fit to the originals that can be derived with this smaller number of factors: e.g., if we take $k = 1$, we shall reach the closest approximation obtainable in terms of a single 'general factor' only.†

The matrix expression‡ of this analysis is sufficiently familiar: but the method that I propose pushes the analysis one stage further

Recent illustrations of work in our laboratory on correlations between persons, using both 'method a' and 'method b', have been described in this journal by my colleague Dr. Stephenson (6); a review of earlier applications of the device, with a discussion of its value and limitations, is given in my own paper on the subject (2).

* The proof of the least squares formula for the saturation coefficients is given in my memorandum for the International Institute Examinations Inquiry (1, pp. 247, 286); the proof of the formula for the multiple correlation coefficient is given by Kelley (9, p. 296). We have only to substitute r_{gj} for Kelley's r_{oj} (i.e., to identify the 'criterion' with the 'general factor', taken as the best weighted sum of the tests themselves), and we are led immediately to equation (6).

† In passing it may be noted that if the 'general factor' is conceived as distributing the individuals, not along a linear scale, but into two discrete classes (such as Dr. Stephenson's 'psychological types', which are described as being 'as discontinuous as the two sexes'), the result is essentially the same: we reach a factor that yields the widest discrimination, no longer between the different individuals, but between the different types: (in the notation used below, the usual regression coefficients, $\mathbf{w} = \mathbf{f}' R^{-1}$, become $\mathbf{w} = \mathbf{d}' R^{-1}$, where the vector $\mathbf{d}$ denotes the differences between the means of the two functions. Except that one determinant has here to be evaluated instead of two, the expression is analogous to that reached by Fisher (6) for somewhat similar problems in other statistical fields).

Where the classification into types is itself determined on the basis of independent multiple measurements, and the problem is to correlate a mental classification with a physical (as in verifying Kretschmer's theory), the issue becomes more complex. As I have indicated elsewhere (21, p. 186), we are then correlating, not a scalar variable with a vector, but one vector with another; and the problem becomes a problem in what may be called bi-multiple (or vector) correlation.

‡ In what follows, the same letter of the alphabet is used to denote a matrix, its vectors (i.e., its rows or columns), and its elements: the matrix will be indicated by a capital letter, the vector by a lower case letter in bold type, and the elements by a lower case letter in italics with a double subscript. So far as possible, I have substituted Thurstone's notation for my own. His book (7) unfortunately had not appeared when my memorandum for the International Examinations Committee was drawn up: further, it seemed at that time desirable that the notation should, so far as possible, preserve the symbols most familiar to English students, namely, those connected with Spearman's methods of analysis.

back than its usual formulation.* The fundamental postulate, and the final outcome, of factor analysis is usually written

$$S = F\,P \tag{1}$$

where $S \equiv$ the 'score matrix' of observed measurements, $F \equiv$ the 'factorial matrix' of factor-loadings or saturation coefficients, $P \equiv$ the orthogonal or semi-orthogonal 'population matrix' giving the hypothetical measurements of the population for the several independent factors as above defined.

Let us, however, put

$$\sum f^2_{i1} = v_1 , \cdots , \sum f^2_{in} = v_n , \tag{2}$$

so that v_j denotes† the amount contributed to the total variance by the j^{th} factor. We may then re-write (1) in the more convenient form

$$S = L\,V^{\frac{1}{2}}\,P , \tag{3}$$

where L denotes the orthogonal matrix of direction cosines specifying the relation of factor-axes to test-axes, and V the diagonal matrix of factor variances.

From this 'canonical resolution' of the score matrix S we at once obtain

$$R = S\,S' = F\,F' \tag{4}$$

$$= L\,V\,L' , \tag{5}$$

where R denotes what in the analysis of variance would be termed the matrix of 'sums of squares and products' and (if the scores are in unitary standard measure) is simply the matrix of correlations between the various tests. It will be seen that R must be real, square, symmetric, positive-definite,‡ and of rank n.

We have now reduced the correlation matrix to its simplest or 'canonical' form. Equations (3) and (5) which express these canonical transformations I regard as fundamental. It is on them that most of my arguments will be based.

In theory the values required for V and L can be obtained by solving the equation

* My equation (1) is Thurstone's equation (2), (7, p. 54); my equation (4) is equivalent to his equations (22) and (39). The general relations between Kelley's analysis and Thurstone's may be expressed by saying that, for Thurstone, $L\,V^{\frac{1}{2}}$ ($\equiv F$) would denote the regression coefficients for determining S from P; for Kelley, L would denote the regression coefficients for determining S from $V^{\frac{1}{2}}P$.

† I have substituted V and v for the Greek letters Λ and λ that are almost universally used for the 'latent' or (as they were once called) 'lambdaic roots' in matrix algebra. The former are the natural symbols for variances in statistical notation: the latter present minor difficulties in typing and in teaching.

‡ i.e., all its principal minors are greater than or equal to zero.

$$R\mathbf{l} = v\mathbf{l}. \tag{6}$$

Here R is a given square matrix, defined as above, and $\mathbf{l}$ a column vector of n elements, as yet undetermined. We obtain, as the condition of a non-trivial solution, n values for v; and, on substituting in equation (6) each of these values in turn, we may solve the n homogeneous equations for the ratios

$$l_{1j} : l_{2j} : \cdots : l_{nj}.$$

If we add the condition that

$$l^2_{1j} + l^2_{2j} + \cdots + l^2_{nj} = 1,$$

the resulting values for l_{kj} may be regarded as direction cosines specifying the direction of the vector $\mathbf{l}_j$ associated with the root v_j. Now, since R is symmetric (i.e., $R = R'$), the roots v must be real, and the vectors $\mathbf{l}$ mutually orthogonal: for

$$R\,\mathbf{l}_i = v_i\,\mathbf{l}_i,$$

and

$$R\,\mathbf{l}_j = v_j\,\mathbf{l}_j;$$

hence,

$$\mathbf{l}'_i\,v_j\,\mathbf{l}_j = \mathbf{l}'_i\,R\,\mathbf{l}_j = \mathbf{l}'_i\,R'\,\mathbf{l}_j = (R\,\mathbf{l}_i)'\,\mathbf{l}_j = (v_i\,\mathbf{l}_i)'\,\mathbf{l}_j = \mathbf{l}'_i\,v'_i\,\mathbf{l}_j;$$

that is,

$$(v_j - v'_i)\mathbf{l}'_i\,\mathbf{l}_j = 0.$$

Accordingly, if $i = j$, $v_j = v'_j$: from which it follows that v_j and its conjugate v'_j (both being scalars) are real and positive.* Again if $i \neq j$, $\mathbf{l}'_i\,\mathbf{l}_j = 0$ (unless the two latent roots happen to be equal). We thus have

$$\sum_k l_{ki}\,l_{kj} = \begin{Bmatrix} 0, \text{ if } i \neq j \\ 1, \text{ if } i = j \end{Bmatrix} \tag{7}$$

Since the values v_j are obtained by solving what is known as the *équation caractéristique* of R, they are called its 'characteristic' or 'latent roots', and in quantum theory its *Eigenwerten*† ('characteristic' or 'proper' values); here, as we have seen, they are to be identified with the factor-variances. The set of values l_{kj}, corresponding to a particular latent root v_j, is termed the 'characteristic' or 'latent vector', and in quantum theory the *Eigenvektor* (the 'principal' or

* The meaning of v'_j will be clearer, if we remove the relevant restrictions. The conclusions in the text only follow as these restrictions are reintroduced.

† There is a tendency among recent German writers (10, pp. 15, 23) to designate the latent roots 'charakterische Zahlen' and to keep the term 'Eigenwerte' for the reciprocals. (For the relations between the two, see 2, pp. 78-79).

'unit proper' vector): here, as we have seen, such sets may be identified with the direction cosines for rotating the test-axes into coincidence with the factor-axes, and their elements are therefore proportional to the corresponding saturation and regression coefficients contained in the factor matrix F.

Once we have made these identifications, the problem of factor analysis in psychology is seen to be precisely analogous to that of reducing a symmetrical matrix to its canonical form in matrix algebra and of determining the 'eigen values' and the 'eigen states' of a real observable in quantum theory.* Incidentally, it may be noted that the results thus reached are independent of all reference to the mode in which the frequencies of the variables are distributed: no assumption of normality has been necessary. Thus, when the algebraic results are expressed in geometrical terms, the ellipsoids obtained by treating the measurements as samples of continuous variables may be regarded simply as 'strain ellipsoids', indicating the relations that arise from changing the scale of measurement differently in different directions, and not as the frequency ellipsoids arising from a normal distribution. At the same time, if problems of probability arise (as where the errors of prediction, or the sum of their squares, are to be minimized), the frequency-interpretation is still available, and the method at once gives an appropriate answer.

Let us now write H_j for the matrix of rank one obtained by multiplying the single column vector $\mathbf{f}_j$ by its transpose, and similarly let let us write E_j for the matrix obtained by multiplying the latent column vector $\mathbf{l}_j$ by its transpose: thus

$$E_j = \mathbf{l}_j \mathbf{l}'_j$$

$$H_j = \mathbf{f}_j \mathbf{f}'_j$$

and

$$H_j = v_j E_j .$$

Then the final expansion of R may be expressed

$$R = H_1 + H_2 + \cdots + H_n \tag{8}$$

$$= v_1 E_1 + v_2 E_2 + \cdots + v_n E_n . \tag{9}$$

Here the H's represent a series of 'hierarchies' (in Spearman's sense);

* In quantum mechanics, complex numbers are used, and the matrices are Hermitian, whereas in factor-analysis the elements in the matrices are all real numbers (though in my view many problems might receive a more general solution if complex numbers were introduced). Nevertheless, since the field of real numbers may be legitimately treated as part of the field of complex numbers, the real axisymmetric matrices, formed by the correlation and covariance tables of psychology, may be regarded simply as special cases of Hermitian matrices, in which the imaginary part of the complex number vanishes.

and, in view of the properties that I shall presently demonstrate, the E's may be termed 'latent' or 'unit hierarchies'. The final expression given by equation (9) may be termed the *canonical expansion of the correlation matrix.* I thus conceive the fundamental problem of factor-analysis to consist in expanding R in terms of its latent roots and latent hierarchies in accordance with this equation. The reduction more usually made — to 'manifest hierarchies', as they might be called — in accordance with equations (4) and (8), seems only an incomplete performance of the essential task.

III. The Reduced Hierarchy and its Properties.

These 'unit' or 'latent hierarchies' have a number of peculiar properties; and these properties in turn will not only reveal more explicitly the striking analogies to which I have alluded, but also facilitate the deduction of useful formulae both for practical and theoretical work.

(i). By definition each matrix E is constructed by post-multiplying the column vector $\mathbf{l}$ by its transpose. It is therefore (like R and H) symmetric: i.e.,

$$E' = (\mathbf{l}\mathbf{l}')' = \mathbf{l}\mathbf{l}' = E. \tag{10}$$

(ii). Since the vector $\mathbf{l}$ is by hypothesis a matrix of one column only and therefore of rank one, it follows that E (like H) is a matrix of rank one; all its rows are linearly dependent, and the whole of the n^2 elements can be deduced from a set of n values only. In explicit notation, since

$$e_{ih} = l_i\, l_h, \tag{11}$$

$$\frac{e_{ih}}{e_{jh}} = \frac{l_i\, l_h}{l_j\, l_h} = \frac{l_i\, l_k}{l_j\, l_k} = \frac{e_{ik}}{e_{jk}} \tag{12}$$

$$(i, j, h, k, = 1, 2, \cdots, n)$$

E therefore obeys the tetrad difference criterion; is strictly hierarchical; and, being singular can possess no inverse.

(iii). It follows, too, that E cannot be expressed as the sum of two other hierarchies (say $F = \mathbf{f}\,\mathbf{f}'$ and $G = \mathbf{g}\,\mathbf{g}'$) unless both these hierarchies are themselves multiples of E : for, consider any tetrad such as

$$\begin{bmatrix} f^2{}_1 + g^2{}_1 & , f_1 f_2 + g_1 g_2 \\ f_1 f_2 + g_1 g_2 , & f^2{}_2 + g^2{}_2 \end{bmatrix} = \begin{bmatrix} l^2{}_1 & , l_1 l_2 \\ l_1 l_2 , & l^2{}_2 \end{bmatrix}$$

Write

$$\frac{f_2}{f_1} = p\,, \text{ and } \frac{g_2}{g_1} = q\,.$$

Then

$$\frac{f^2{}_1 + g^2{}_1}{p\,f^2{}_1 + q\,g^2{}_1} = \frac{p\,f^2{}_1 + q\,g^2{}_1}{p^2\,f^2{}_1 + q^2\,g^2{}_1}\,;$$

i.e.,

$$(p^2 + q^2)\,f^2{}_1\,g^2{}_1 = 2\,p\,q\,f^2{}_1\,g^2{}_1\,.$$

Hence

$$p = q\,,$$

and

$$\frac{f_2}{f_1} = \frac{g_2}{g_1} = \frac{l_2}{l_1}\,. \tag{13}$$

Similar relations hold for all the other elements in the vectors f and g. Equation (11) thus implies that the matrices E_j are ultimate or 'pure'; i.e., they cannot themselves be further decomposed, as R has been decomposed into the sum of the E's.

(iv). From (11) we have

$$\sum_i e_{ih} = l_h \sum_i l_i\,. \tag{14}$$

Thus the sums of the columns (or of the rows) of any given matrix E are (like the elements in each column or row itself) proportional to the elements of the latent vector from which E has been constructed.

The total of the elements in E is therefore

$$\sum_h \sum_i e_{ih} = (\sum_i l_i)^2\,. \tag{15}$$

And summing these totals for all the n E-matrices we have

$$\begin{aligned} &(\sum_j l_{i1})^2 + (\sum_j l_{i2})^2 + \cdots + (\sum_j l_{in})^2 \\ &= \sum_j l^2{}_{1j} + \sum_j l^2{}_{2j} + \cdots + \sum_j l^2{}_{nj} \\ &= n\,. \end{aligned} \tag{16}$$

In matrix notation the result may be reached more simply as follows:

$$\begin{aligned} &E_1 + E_2 + \cdots + E_n \\ &= \mathbf{l}_1\,\mathbf{l}'_1 + \mathbf{l}_2\,\mathbf{l}'_2 + \cdots + \mathbf{l}_n\,\mathbf{l}'_n \\ &= LL' \\ &= I. \end{aligned} \tag{17}$$

These results (equations 15, 16, 17) will incidentally provide a useful means of checking the arithmetical calculations.

(v). Since $L'L$ also equals I, we have

$$\sum_i l^2_{i1} = \sum_i l^2_{i2} = \cdots = \sum_i l^2_{in} = 1 \; ; \tag{18}$$

i.e., the sum of the diagonals in any matrix E_j (its 'trace') is equal to unity. These properties may be conveniently summed up by describing E as a 'unit hierarchy'.

(vi). A still more peculiar property is that

$$E^2_j = (\mathbf{l}_j \mathbf{l}'_j)(\mathbf{l}_j \mathbf{l}'_j) = \mathbf{l}_j \mathbf{l}'_j = E_j ,$$

and generally

$$E^m_j = E_j \; ; \tag{19}$$

that is, each latent hierarchy is 'idempotent'.

(vii). Again

$$E_i E_j = (\mathbf{l}_i \mathbf{l}'_i)(\mathbf{l}_j \mathbf{l}'_j) = 0 \; , \qquad (i \neq j) \; ; \tag{20}$$

that is, any two latent hierarchies derived from the same expansion are mutually orthogonal.

(viii). The latent hierarchies, however, may be derived without reference to the latent vectors: for, on examining the form of the characteristic matrix, it is easily seen that E_1 (say)

$$= \frac{(R - v_2 I)(R - v_3 I) \cdots (R - v_n I)}{(v_1 - v_2)(v_1 - v_3) \cdots (v_1 - v_n)} , \tag{21}$$

and similarly for all the matrices E_j.

(ix). Finally, it may be of interest to consider how the variance for a given factor changes with changes in the several correlations. From equation (5) we have $V = L' R L$; hence each v_j can be obtained as a sum of terms such as $\sum_i \sum_k l_{ij}\, l_{kj}\, r_{ik}$. By partial differentiation for each particular correlation, r_{ik}, we have

$$\frac{\partial v_j}{\partial r_{ik}} = l_{ij}\, l_{kj} \, . \tag{22}$$

Thus E_j proves to be the result of a matrix differential operator whose ik^{th} (and ki^{th}) element is $\dfrac{\partial}{\partial r_{ik}}$.

Equations (21) and (22) show that what we have been calling the latent hierarchies of the matrix R are simply the numerators of

its 'partial resolvents'.* Equations (19) and (20) further indicate that each E_j has all the essential properties of what, in the theory of groups, is termed a 'selective operator', or (when the groups can be represented by linear transformations) a 'projective operator', i.e., an operator which projects any given vector on to the unit vector $\mathbf{l}_j$ defined by the corresponding latent root v_j .† A set of selective operators satisfying equation (17) as well as (19) and (20) is called a 'spectral set'; 'it analyzes a mixed aggregate into pure constituents as light is analyzed by a prism or grating into a spectrum of pure colours'.‡ Including (13) we may say that between them the four equations secure that the set of operators which represent the several factors shall be (a) exhaustive, (b) non-overlapping, (c) pure (i.e., indecomposable), and (d) idempotent (i.e., such as to yield the same result no matter how many times the process of 'selection' is applied

* The 'resolvent' of R is the reciprocal of $\lambda I - A$, where λ may have any value other than the latent roots: (cf. Cullis, 11, pp. 316 *et seq.*, Turnbull & Aitken, 12, pp. 160, 163, 184).

† For the formal definition of a projective operator see Weyl, (13, p. 23), or Neumann, (14, pp. 39-41) and Satz, (12), and, for a fuller and more technical discussion, Stone, (15, chapter IV, on 'Resolvents, Spectra, and Reducibility'). I know of no discussion bringing together *all* the properties enumerated above.

‡ The name has arisen from the fact that in the realm of atomic physics the most simple and direct experimental illustrations of the 'laws of measurement' — laws expressed by theorems very similar to those set out above — is to be found in the spectral analysis of molecular rays along the lines of the classical experiments of Stern and Gerlach (13). In quantum theory a 'complete observation' of a given mixed or inhomogeneous aggregate or assembly is regarded as a species of spectral analysis in which the given aggregate is resolved into a number of constituent parts which are relatively pure or homogeneous with respect to some particular variable. At first sight, the physicist's problem as thus stated seems closely analogous to that of observing and measuring sample assemblies of mental traits in sample assemblies of the population with a view to discovering the distribution of relatively pure factors; and the more recent extension of the spectral theory to continuous as distinct from discrete spectra is reminiscent of the problems that arise in considering the continuity or discontinuity of the fundamental mental traits. There are, of course, differences of aim as well as resemblances (e.g., in mental testing we are generally more interested in measuring the individual than the assembly); but into these it is impossible to enter here.

I may, however, point out that the analogy goes deeper still. Of late a good deal of interest has been aroused by the 'phenomena of constancy' in all processes of perception (cf. Katz, 17). Now, if we recognize that, in psychology as in modern physics, an 'observation' is to be specified not by a quantity but by a structure, not by the correspondence between a single sensation and a single stimulus, but by the correspondence between patterns or matrices, then we may regard the 'real object' as specified by the canonical form of that matrix (containing the Eigenwerte) which is implicitly thought of as providing a constant and fundamental standard — as in fact a kind of Platonic εἶδος. As I walk round my study table, I am aware, not of the varying two-dimensional versions of it that reach my retina in changing perspective, but a kind of standardized four-square table in three-dimensional space. Now if, instead of the isolated stimuli, the entire matrix of stimuli is treated as the unit, the transformations of the pattern are quite as simple to examine as the transformations of the isolated stimulus, and the inversion of such transformations simpler still. It may be added that the postulate that introspection shall not distort the state observed is expressed by the requirement that in such cases the observation shall be represented by an idempotent operator.

and reapplied). Or, to adopt the convenient language of geometry, we may say that, in factor analysis as in quantum theory, the introduction of these projective operators enables us to resolve "the total representational n-dimensional space" (i.e., what in correlational work has been called the "common factor space" (Thurstone, 7, p. 69)) into n "characteristic linear subspaces", each representing an independent factor.* The success which has attended this method in recent developments of physics is, I think, a guarantee of its value in solving many corresponding problems in psychology.

IV. Corollaries for Practical Work.

Apart from the theoretical suggestiveness of this result, the practical properties just enumerated yield several important results that directly bear on the practical task of arithmetical evaluation. To determine v_j and E_j, the direct method, as we have seen, would be to solve the characteristic equation for the v_j and then substitute the values for the v_j in turn and solve for the $\mathbf{l}_j$. If, however, the number of traits or tests is large, the labour of computing the determinants is prohibitive. It is for this reason, I take it, that so many different 'methods of factor analysis' have been proposed: these claim to give, with relatively little trouble, not an exact, but a sufficiently approximate result. Now I have elsewhere (2) endeavoured to show that, by expanding the higher powers of R in terms of its latent hierarchies,

* Cf. Weyl, (11, p. 22). This mode of formulation enables us to express many other subsidiary problems that arise in psychological analysis in a shape identical with that of problems that have already been solved in the field of quantum-mechanics. One important problem is to ascertain whether the same set of factors is common to different sets of tests (with the same tested persons) or to different samples of the population (with the same set of tests). If S_1, S_2, R_1 and R_2 represent the scores and correlations from the two sets, then, when the factors (and therefore their direction cosines L) are the same, we have at once $R_1 R_2 = R_2 R_1$, and (since the factor measurements in P_1 and P_2 will tend to be uncorrelated) $S_1 S_2' = S_2 S_1'$. This 'symmetry criterion' has already been applied with success by Miss Williams and Miss Davies in the researches on personal types. Again, another problem is to ascertain the minimum number of independent factors requisite to account for S and R within the limits imposed by the probable error: this also receives a simple solution (see below). More particularly, in quantum theory R is said to be 'completely reducible' if only two sub-spaces are required: this situation is analogous to that postulated by the bifactor hypothesis in psychology; and the conditions deduced in physical theory can be carried over to psychological theory. Again, to allow for 'chance' factors and for the 'errors' of observation or measurement, space of infinite dimensions will in the end be required: but it has been shown in quantum theory that, in general, theorems proved for finite matrices may be assumed true for infinite, and the recent extensions of the theory have revealed both the justifications and the limitations of such an assumption (cf. Hilbert and Courant, 10, pp. 128-9). Once more, the treatment of issues involving the determination of probabilities ('laws of transition' and 'exchange relations') could be adapted with but little modification.

one may obtain, not merely a reconciliation of the rival methods so far suggested, but also a simple modification which will rapidly yield a result of any desired exactitude.

If we begin by squaring the covariance matrix R, we have

$$
\begin{aligned}
R^2 &= (v_1 E_1 + v_2 E_2 + \cdots + v_n E_n)^2 \\
&= v^2_1 E^2_1 + v^2_2 E^2_2 + \cdots + v^2_n E_n \\
&+ 2\, v_1 v_2 E_1 E_2 + 2\, v_1 v_3 E_1 E_3 + \cdots + 2\, v_{n-1} v_n E_{n-1} E_n \\
&= v^2_1 E_1 + v^2_2 E_2 + \cdots + v^2_n E_n \qquad (23)
\end{aligned}
$$

since by equation (19) $E^2_1 = E_1, \cdots, E^2_n = E_n$, and by equation (20) $E_1 E_2 = E_1 E_3 = \cdots = E_{n-1} E_n = 0$. On continuing the self-multiplication of the matrix, we shall evidently obtain for any* power of R:

$$R^m = v^m_1 E_1 + v^m_2 E_2 + \cdots + v^m_n E_n . \qquad (24)$$

Dividing through by v^m_1 we have

$$R^m/v^m_1 = E_1 + v^m_2/v^m_1 E_2 + \cdots + v^m_n/v^m_1 E_2 \ ;$$

and, since $v_1 > v_2 > \cdots > v_n$, we can, by taking m large enough, obtain

$$R^m/v^m_1 = E_1 \qquad (25)$$

to as close an approximation as we may desire. Similarly, by taking $(m - 1)$ large enough, we can obtain

$$R^m/v_1^{m-1} = v_1 E_1 = H_1 ,$$

that is,

$$R^m = v_1^{m-1} H_1 . \qquad (26)$$

It is thus evident that, with a sufficient number of self-multiplications, any symmetrical matrix, such as a table of correlations or covariances, *can be reduced as closely as we wish to a matrix of rank one*, i.e., to a Spearman hierarchy. To factorize such a matrix is now exceedingly simple: for, since its rank is only one, it contains but one common factor.

To determine the saturation coefficients which form the vector $\mathbf{l}_1$, we may apply equation (14). Their proportionate values can thus be obtained by simply adding the columns of the product matrix R^m; and on normalizing the proportionate values we reach $\mathbf{l}_1$.

Since $R = S\, S'$ (where S is the matrix of scores), the elements

* This includes negative indices: so that we have here a rapid means of calculating the inverse of R, which is often wanted for the regression equations. Note also, that, since

$$R\, R^{-1} = E_1 + E_2 + \cdots + E_n, \text{ and } R\, R^{-1} = I ,$$

we have a simplified proof of equation (17).

of R may be termed 'product-moments of the first order'; those of R^m may accordingly be termed 'product-moments of the m^{th} order'. When $v_2, \cdots, v_n$ are absolutely as well as relatively negligible, i.e., when R itself is virtually hierarchical, the summation formula may be applied to R as it stands:* otherwise it must be applied to 'higher product moments' instead of to the observed correlations: i.e., instead of

$$\mathbf{f}'_0 R = c_1 \mathbf{f}'_1 \text{ ('simple summation')} \tag{28}$$

we may take

$$\mathbf{f}'_0 R^m = c_m \mathbf{f}'_m = c_m \mathbf{f}'_{00} \tag{29}$$

where $\mathbf{f}_j$ denotes the j^{th} approximation to the saturation coefficients for the first factor; $\mathbf{f}_0$ denotes the summation operator $\{1, 1, \cdots, 1\}$; $\mathbf{f}_{00}$ denotes the true value of the saturation coefficients, i. e., the vector $\{r_{1g}, \cdots, r_{ng}\}$, and can be determined as exactly as we desire by taking m sufficiently large; and c_j denotes a constant that can be explicitly determined if required.

Now the summation of a power of R (R^4, say) may be expressed $\mathbf{f}'_0(R \times R \times R \times R)$. But this is clearly equal to $\mathbf{f}'_0 R(R \times R \times R)$. It follows that $\mathbf{f}'_0 R^m$ may be computed in two ways — either by table-by-table multiplication or by column-by-table multiplication. Thus, we may begin by multiplying the matrix by itself, repeat the multiplications again and again to the m^{th} factor, and then add the columns of the final product; or we may begin by adding the columns of the initial table, then use these sums to weight the coefficients in each row, and add the columns again, repeating the additions again and again to the m^{th} sum. Incidentally, since $\mathbf{w}_0(R \times R \times R \times R) = \mathbf{w}_0(R^2)^2$, we may conveniently abridge the first method by taking m to be a power of 2, say 2^p; for, if we square the product of each squaring, we shall make p table-by-table multiplications furnish the same result as m.

* In an early paper (18, 1917) I showed that, if R is hierarchical, the saturation coefficients may readily be determined by the 'simple summation formula'

$$r_{ig} = \sum_j r_{ij} / \sqrt{\sum_i \sum_j r_{ij}}. \tag{27}$$

But I have always maintained that, so far as R is *not* hierarchical, this formula could be regarded only as a first approximation. Thurstone, however, (7, p. 94, equation (13)), has more recently taken the formula as the basis of his centroid method, and treated it as applicable to any form of correlation table. I myself have always maintained that, so far as R is not hierarchical, the formula could be regarded as furnishing a first approximation only; and in another article I have endeavoured to demonstrate at length, and with arithmetical illustrations, that the saturation coefficients or 'factor loadings' obtained by the centroid method are, in effect, simply the first of a converging sequence of values of which the limits are the values given by the direct solution of the characteristic equation.

We have, therefore, a choice between a large number of short multiplications and a small number of long multiplications. As a rough rule, it may be said that table-by-table multiplication will prove more economical when the table is small, and may also be of service in the reduction of highly irregular tables: whereas table-by-column multiplication will prove quicker when the table is large, and may also be of service in the final stage of the whole calculation. Both these methods rest on the same principles as the procedure proposed by Hotelling; and variants of them are familiar to computers in other mathematical fields.

Hotelling's original iterative method (19) involves a series of table-by-column multiplications; and more recently (20) he has suggested the matrix-squaring device as a useful preliminary to iterations with the table-by-column method. As my students and I have applied them, however, the principles described above entail several minor but important differences, which (as it appears to us) at once diminish the labour and enhance the accuracy of the final results. Hotelling's instructions are as follows: "The process should be started with trial values of one digit each . . . as judged by inspection. Each digit should be accurately determined by repetition until stationary values are reached before the calculations are carried to another place" (19, pp. 431-2; cf. 20, p. 33). With the table-by-column method, the procedure* required by the foregoing proof is as follows: (i) instead of more or less arbitrary trial-values, judged by inspection, we commence always with the figures obtained by simple summation, and use these as the first set of multipliers or weights; (ii) at every stage we treat the sums not as suggesting new 'trial values' to be chosen by further comparison or an intelligent guess, but as precise weights, working with the full number of significant digits from the start: this renders the procedure perfectly mechanical at every step, and does not throw upon the inexperienced calculator the task of judging what corrections he shall try in the hope of jumping at once to a close approximation; (iii) when the full figures are retained, then, as we keep multiplying, the way in which the product sums are approximating to the final values becomes obvious and it is seen that the differences are apparently diminishing according to a regular rule: hence, after three or four iterations have been calculated explicitly, the results of the subsequent iterations may be obtained by the method of differences; and, in most cases, the final values may be obtained forthwith by extrapolation.

So far as theory is concerned, instead of regarding the table-by-table method as a modification of or a preliminary to the table-by-column method, I regard the principle of self-multiplication of the entire table as fundamental. In either case, however, the point to stress is that we are not dealing merely with alternative modes of approximation, but with alternative procedures that lead to identical results. This has been so often overlooked or questioned† that I may be pardoned for giving a concrete example. I take the table used to illustrate the two procedures in my previous article, namely, that constructed by Kelley to compare his own method with that of Thurstone.

Σr_{ij}	r_{i1}	$r_{i1}\Sigma r_{ij}$	r_{i2}	$r_{i2}\Sigma r_{ij}$	r_{i3}	$r_{i3}\Sigma r_{ij}$
1.96	1.00	1.9600	.70	1.3720	.26	.5096
1.90	.70	1.3300	.75	1.4250	.45	.8550
1.06	.26	.2756	.45	.4770	.35	.3710
Σr_{ij} and $\Sigma\Sigma r_{ij} r_{ij}$	1.96	3.5656	1.90	3.2740	1.06	1.7356

* For illustrations cf. (2, p. 91); (3, pp. 185-8, Tables IV and V).

† An instance in which this principle has apparently been overlooked is the use of the simple summation formula (the so-called centroid formula, described above, equation 27) to determine the saturation coefficients. Here the sum of the correlations for each test must be the same as the correlations for each test with the sum. Hence, as we have seen, the saturation coefficients thus computed simply treat the general factor as identical with the simple sum or average of the test scores (1, p. 287, footnote 1).

The totals $\Sigma\, \Sigma r_{ij}\, r_{ij}$ form the vector $(\mathbf{f}_0\, R) \times R = \mathbf{f}_0\, R^2$. If now we square R first of all and then add, we obtain

	$\Sigma r_{1j}\, r_{ij}$	$\Sigma r_{2j}\, r_{ij}$	$\Sigma r_{3j}\, r_{ij}$
	1.5576	1.3420	.6660
	1.3420	1.2550	.6770
	.6660	.6770	.3926
$\Sigma\, \Sigma\, r_{ij}\, r_{ij}$	3.5656	3.2740	1.7356

Here the totals $\Sigma\, \Sigma r_{ij}\, r_{ij}$ form the vector $\mathbf{f}_0(R \times R) = \mathbf{f}_0\, R^2$. It will be noted that no decimal figures have been omitted, and the results are exact.

When $\mathbf{f}$ has been thus determined as accurately as is required, we shall have $\mathbf{f}_{m-1} = \mathbf{f}_m$ to the number of digits retained, and

$$\mathbf{f}'_{m-1}\, R^m = v^m{}_1\, \mathbf{f}_m\, .$$

To evaluate v_1, two methods are available.

(i) Since, within the limits of our approximation, $R^m = v^m{}_1\, E_1$, and since by equation (18) the sum of the diagonals in $E_1 = 1$, it follows that the sum of the diagonals in $R^m = v^m{}_1$. (30)

(ii) Since, too, $R^{m-1} = v_1{}^{m-1}\, E_1$
we have

$$v_1 = \frac{\mathbf{f}'_0\, R^m\, \mathbf{f}_0}{\mathbf{f}'_0\, R^{m-1}\, \mathbf{f}_0}$$

$$= \frac{\text{Total of all coefficients in } R^m \text{ (i.e., of weighted sums)}}{\text{Total of all coefficients in } R^{m-1} \text{ (i.e., of weights)}}\, . \qquad (31)$$

To keep c constant, and so render the progress of the approximations visible to the eye, we may at any stage reduce consecutive sets of product-sums to the same order by dividing them by the ratio given above (31): or, better still, we may from the very start express each set of weights as fractions of unity in the usual way; in this case, each set of product-sums will be first divided by its own total before being used to re-weight the original correlation matrix.*

How to determine the constant c and to ascertain the amount of error involved at any stage of the approximation will become obvious if we set down the entire operation in matrix notation, and apply the expansion given above (equation 24). After the m^{th} summation, the vector of weights — i.e., of product-sums when reduced to fractions

* I need not give further illustrative tables: in my review (3) of Kelley's book I have taken his own set of correlations by way of example; equation 30 is exemplified by the table on p. 185, and equation 31 by the table on p. 188.

of their own total — will be

$$\mathbf{w}_m = \frac{\mathbf{f}'_0 R^m}{\mathbf{f}'_0 R^m \mathbf{f}_0}$$

$$= \frac{\mathbf{f}'_0(v^m{}_1 E_1 + v^m{}_2 E_2 + \cdots + v^m{}_n E_n)}{\mathbf{f}'_0(v^m{}_1 E_1 + v^m{}_2 E_2 + \cdots + v^m{}_n E_n)\mathbf{f}_0}$$

$$= \frac{v^m{}_1 \mathbf{t}_1 + v^m{}_2 \mathbf{t}_2 + \cdots + v^m{}_n \mathbf{t}_n}{v^m{}_1 T_1 + v^m{}_2 T_2 + \cdots + v^m{}_n T_n}$$

where $\mathbf{t}_j$ (a vector) is used to denote the sums of the columns in E_j, and T_j (a scalar) to denote the grand total of these sums. On dividing out this yields (retaining only the largest terms)

$$\frac{\mathbf{t}_1}{T_1} - \frac{v^m{}_2}{v^m{}_1} \cdot \frac{(\mathbf{t}_1 T_2 - \mathbf{t}_2 T_1)}{T^2{}_1} - \frac{v^m{}_3}{v^m{}_1} \cdot \frac{(\mathbf{t}_1 T_3 - \mathbf{t}_3 T_1)}{T^2{}_1}$$

$$+ \frac{v_2{}^{2m} T_2}{v_1{}^{2m} T_1} \cdot \frac{(\mathbf{t}_1 T_2 - \mathbf{t}_2 T_1)}{T^2{}_1} + \cdots. \tag{32}$$

The first term is the vector $r_{1g}/\sum r_{jg}$, $r_{2g}/\sum r_{jg}$ $\cdots$; and c therefore has now become $1/\sum r_{jg}$. The second and third terms in (32) indicate the chief sources of error, and therefore the approximate rate at which the inaccuracy is diminishing. If $v^m{}_2/v^m{}_1$ is small, $v^m{}_3/v^m{}_1$ will usually be far smaller, and $v_2{}^{2m}/v_1{}^{2m}$ wholly negligible. Generally, therefore, the amount of error incurred by taking R^m in place of R^∞ will be nearly proportional to $v^m{}_2/v^m{}_1$, and thus virtually diminish in geometrical progression. This is the justification for estimating the later terms of the series $\mathbf{f}_0 R^m$ (as suggested above) by extrapolation instead of by continued iteration.* Having obtained figures as exact as are desired, the simplest method of eliminating c is to normalize the results, thus obtaining the direction cosines or 'regression coefficients'. The saturation coefficients can then be obtained by multiplying the normalized figures by $v^{\frac{1}{2}}$.

* The obvious device is to use the well known formula for the sum of an infinite diminishing geometrical progression: but with a machine continuous multiplication by the constant ratio is almost as quick. It may be noted that the 'common ratio' obtained in extrapolating for the first factor enables us to estimate at once the size of the variance for the second factor. Hence, without further calculations, we may decide whether the variance contributed by this second factor is also large enough to be statistically significant. Working instructions, with examples, have been drawn up for students who are unable to follow the algebraic demonstration but wish to apply the methods here described, and have been in use at the laboratory: they are too lengthy to print in an article, but may be obtained from the Psychological Laboratory, University College, London. Dewar has compared the results of what we call the "G. P." (geometrical progression) method with those of other procedures in a paper to the British Psychological Society (March, 1937) which will no doubt shortly be published.

After the saturation coefficients for the first factor — $\mathbf{f}_1$ and therefore H_1 — have been thus determined, the residuals $(R - H_1)$ can be calculated in the usual way; and, if significant, the saturation coefficients and hierarchy for the next factor may be computed if required. Similarly for the rest of the n factors, until all the significant terms of the expansion (9) have been evaluated.

To decide whether a single general factor is alone sufficient to account for the observed correlations, i.e., whether $R = H_1$ within the limits of the sampling errors, there is no need to calculate all the second order minors ('tetrad differences') and their probable errors. We have only to inquire whether the residual variance $v_t - v_1 \equiv v_2 + \cdots + v_n$ is statistically significant or not. I may add that, if we construct an artificial matrix of correlations from a given set of independent factors, the foregoing method of analysis appears to be the only one which will lead back to the factor-variances and saturation coefficients as originally given.*

REFERENCES

(1) BURT, C., "Memorandum on the Analysis of Examination Marks," ap. Hartog, P. and Rhodes, E. C. *Marks of Examiners.* London: Macmillan and Co., 1936.

(2) BURT, C., "Correlations between Persons," *Brit. Journ. Psychol.*, 1937, **28**, pp. 56-96.

(3) BURT, C., "Factor Analysis with and without Successive Approximations," *Brit. Journ. Educ. Psychol.*, 1937, **7**, pp. 172-195.

(4) FISHER, R. A. *Statistical Methods for Research Workers*, London: Oliver & Boyd, 1934.

(5) KELLEY, T. L. *The Essential Traits of Mental Life*, Cambridge, Mass.: Harvard University Press, 1935.

(6) STEPHENSON, W. "Introduction to Inverted Factor Analysis," *Journ. Educ. Psychol.*, 1936, **27**, pp. 353-367.

(7) THURSTONE, L. L. *The Vectors of Mind*, Chicago: University of Chicago Press, 1935.

(8) FISHER, R. A. "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, 1936, **7**, pp. 179-188.

(9) KELLEY, T. L. *Statistical Method*, New York: Macmillan, 1923.

(10) COURANT, R. UND HILBERT, D. *Methoden der Mathematischen Physik*, I. 2 te Auflage. Berlin: J. Springer, 1931.

(11) CULLIS, C. E. *Matrices and Determinoids*, vol. III. Cambridge: University Press, 1925.

(12) TURNBULL, H. W. AND AITKEN, A. C. *An Introduction to the Theory of Canonical Matrices.* London: Blackie, 1932.

(13) WEYL, H. *The Theory of Groups and Quantum Mechanics.* London: Methuen, 1931.

(14) NEUMANN, J. V. *Mathematische Grundlagen der Quantenmechanik.* Berlin: J. Springer, 1932.

* I am indebted to Dr. G. F. Trubridge for his kindness in reading my manuscript and drawing my attention to several *corrigenda.*

(15) STONE. *Resolvents, Spectra, and Reducibility*, Chap. IV.
(16) STERN, O. AND GERLACH, W. *Zeitschr. für Physik*, 1921, **7**, pp. 249 *et seq.*; 1922, **8**, pp. 110 *et seq.*
(17) KATZ, D. *The World of Colour.* London: Kegan Paul, 1935.
(18) BURT, C. *Distribution & Relations of Educational Abilities.* London: P. S. King, 1917.
(19) HOTELLING, H. "Analysis of a Complex of Statistical Variables into Principal Components," *Journ. Educ. Psychol.*, 1933, **24**, pp. 417-441, 498-520.
(20) HOTELLING, H. "Simplified Calculation of Principal Components," *Psychometrika*, 1936, **1**, pp. 27-35.
(21) BURT, C. "The Analysis of Temperament," *Brit. J. Med. Psych.*, 1938, **17**, pp. 158-188.

PSYCHOMETRIKA—VOL. 3, NO. 3
SEPTEMBER, 1938

CONTRIBUTION TO THE MATHEMATICAL BIOPHYSICS OF ERROR ELIMINATION*

The theory of error elimination developed by N. Rashevsky is extended and generalized. Theoretical conclusions obtained are compared with the experimental data available.

HERBERT D. LANDAHL
The University of Chicago

In a paper by N. Rashevsky (1), a theory of delayed conditioned reflexes is developed and applied to the theory of error elimination. It is the purpose here to revise and extend the treatment of the problem without changing the physical assumptions regarding the process of conditioning. A somewhat more general discussion of the problem of maze learning will then be outlined qualitatively and the results compared with available experimental data. Several experiments are suggested since they follow directly from a consideration of the problem.

The following equations are developed by N. Rashevsky (1):

$$R = F(1 - e^{-an})\, I\,, \tag{1}$$

$$I(x) = I_0\, e^{-\beta x}\,, \tag{2}$$

$$I(x,n) = I_0\, e^{-\beta x} - b \int_0^{x_0} R(x,n)\, dx\,, \tag{3}$$

where R is the strength of the conditioned response, n the number of trials, I the intensity of the stimulus, x the abscissa of a particular nerve center; a, β and F are parameters depending on the nerve constants, and b is the proportionality constant for inhibition. The following definitions are also given: the time t, the velocity of wave propagation v, and the time between the stimulus and secondary response, or the delay time for conditioning τ. For $t = 0$, $x = x_0$ and for $t = \tau$, $x = 0$.

Let us consider a slightly modified form of the behaviour pattern discussed by N. Rashevsky. Consider the pattern

$$\begin{array}{ccccc} S_0 \rightarrow R_1 & \text{-------} & R_a & \text{-------} & S_1 \rightarrow R_1 \\ 0 & & \tau_0 & & \tau \end{array}$$

* This investigation has been made possible by a grant from the Rockefeller Foundation to the University of Chicago.

The stimulus S_0, occurring at $t = 0$, produces the unconditioned response R_1, which, after a time τ, results in S_1, a stimulus which produces $\overline{R}_1$ unconditionally. Neglect the reaction time between S_1 and $\overline{R}_1$. Let $\overline{R}_1$ be a response which is the negative of R_1. Also let the physical situation be such that if the response R_1 continues past a particular part, R_a, which occurs at the time τ_0, then the response $\overline{R}_1$ is unavoidable.

More specifically, let us assume that S_0 represents the "sight" of an alley entrance, R_1 is the act of starting forward to enter, R_a is the act of entering, and $\overline{R}_1$ is the act of turning back due to the alley end. τ_0 is then the time between the stimulus S_0 and the entering of the alley. If q is the average velocity of travel in an alley, we have

$$\tau = \tau_0 + \frac{l}{q}, \qquad x_0 = v\,\tau_0 + \frac{v\,l}{q}. \tag{4}$$

In equation (3), we have taken the upper limit of the integral to be x_0, thereby assuming that the centers near x_0 are not inhibited by the time elimination of the error occurs, or that the range of inhibition is small. It is evident from the form of the wave equation that the effect of conditioning near x_0 is small compared to the total conditioned response over the range from x_0 to x.

Substituting the value of $R(x,n)$ from equation (1) in (3), we have:

$$I(x,n) = I_0\, e^{-\beta x} - b \int_0^{x_0} F(1 - e^{-an})\, I(x,n)\, dx\,, \tag{5}$$

for which we have the solution (2):

$$I(x,n) = I_0 \left[e^{-\beta x} - \frac{F\,b\,(1 - e^{-an})\,(1 - e^{-\beta x_0})}{\beta[1 + F\,b\,(1 - e^{-an})\,x_0]} \right]. \tag{6}$$

Now since the intensity is taken as sufficiently strong to elicit all the response conditioned, we have for the total response at x, after n trials

$$R_T(x,n) = \int_x^{x_0} R(x,n)\, dx\,, \tag{7}$$

Substituting from equation (6) in (1) and then in (7), we have:

$$R_T(x,n) = \frac{F\,I_0}{\beta}\,(1 - e^{-an})$$
$$\times\left[e^{-\beta x} - e^{-\beta x_0} - \frac{F\,b\,(1 - e^{-an})\,(1 - e^{-\beta x_0})\,(x_0 - x)}{1 + F\,b\,(1 - e^{-an})\,x_0}\right]. \quad (8)$$

Let n_1 be the number of trials for which the conditioned response at the moment $\imath$, is of itself, without the secondary stimulus, sufficient to produce the response $\overline{R}_1$, or $R_T(0,n_1) = \overline{R}_1$. Equation (8) then becomes:

$$\overline{R}_1 = \frac{F\,I_0}{\beta}\,(1 - e^{-an_1})\left[\frac{1 - e^{-\beta x_0}}{1 + F\,b\,(1 - e^{-an_1})\,x_0}\right]. \quad (9)$$

Solving for n_1, we have

$$n_1 = -\frac{1}{a}\log\left[1 - \frac{\overline{R}_1\,\beta}{F[I_0(1 - e^{-\beta x_0}) - \overline{R}_1\,b\,\beta\,x_0]}\right]. \quad (10)$$

Equations (10) and (4) then give the function $n_1(l)$, the relation between the number of trials for elimination of the alley end and the length of the alley.

If we let

$$\varphi(x_0) = I_0(1 - e^{-\beta x_0}) - \overline{R}_1\,b\,\beta\,x_0 \quad (11)$$

we must have $\varphi_{max} > \overline{R}_1\,\beta/F$ for positive, real values of n_1. For φ to be positive we must have $b < I_0/\overline{R}_1$. The lower and upper limits, l_1 and l_2 of the alley length l, are given by the roots of the equation $\varphi = \overline{R}_1\beta/F$ where we substitute the value for x_0 from equation (4). Now the value of b must be taken very small in order that the range from l_1 to l_2 be large and also in order that the relation $n_1(x_0)$ satisfy more general situations which require that the minimum should occur for values of x_0 which are small compared with the whole range. Then taking b to be very small, φ increases rapidly to its maximum value and thereafter decreases slowly. In this case φ_{max} occurs for a value of l near l_1. Since the function $n_1(l)$ increases monotonically with the reciprocal of φ, the graph of the function may be represented as in Figure 1, and n_1 increases with l except possibly for small values of l.

The value of l_1, the lower limit of l, may be either negative or positive. If l_1 is negative, a finite number of trials is required to eliminate an alley of infinitesimal length. But if l_1 is positive, the alley is not eliminated unless its length exceeds l_1. We might require that the parameters satisfy the condition that l_1 is negative, but this

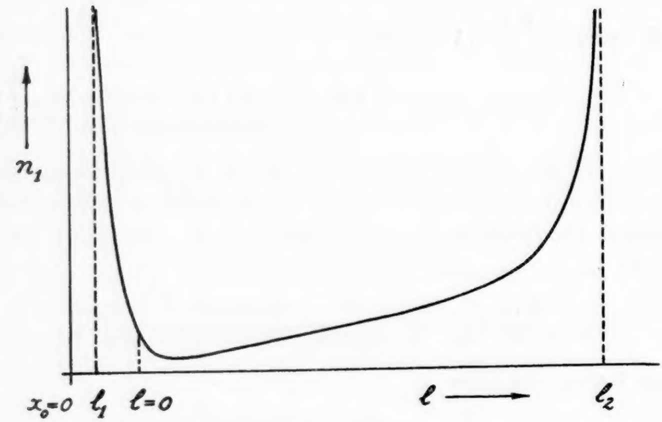


FIGURE 1

is not essential since a sufficiently short alley would never be entered. The upper and lower limits do, however, give some restriction which may assist in the evaluation of the parameters.

Since the upper limit of the integral of equation (3) was taken to be x_0, $R(x,n)$ is always positive, so that $R_T(x,n)$, in equation (7), increases monotonically as x varies from x_0 to zero. Hence the response at $(x_0 - v\,\tau_0)$, the abscissa corresponding to the alley entrance, is less than the response at any point between $(x_0 - v\,\tau_0)$ and zero. There is then no possibility of the elimination of the entrance before the end of the alley. Also, the entrance of the alley will not be eliminated, if the end cannot be eliminated, that is, when l does not lie between l_1 and l_2.

Let us now consider the response for trials after n_1. As the conditioning proceeds, the total response must equal $\overline{R}_1$ at some center $\bar{x}$ greater than zero. Evidently $\bar{x}$ depends on m, the number of trials after n_1 and determines the position at which the animal turns about in the maze. Integrating $R(x,n_1)$ with respect to x from $\bar{x}$ to x_0, we obtain $R_T(\bar{x},n_1)$, the total response due to the conditioning from the first n_1 trials, when the wave front has reached the center $\bar{x}$. Then

$$R_T(\bar{x},n_1) = \frac{F\,I_0}{\beta}\,(1 - e^{-an_1})$$
$$\times\left[e^{-\beta\bar{x}} - e^{-\beta x_0} - \frac{F\,b\,(1 - e^{-an_1})\,(1 - e^{-\beta x_0})\,(x_0 - \bar{x})}{1 + F\,b\,(1 - e^{-an_1})\,x_0}\right]. \quad (12)$$

Substituting the value of n_1 from equation (10) into (12) we have:

$$R_T(\bar{x})_{n_1} = \bar{R}_1 \left[1 - \frac{I_0(1 - e^{-\beta \bar{x}}) - \bar{R}_1\, b\, \beta\, \bar{x}}{I_0(1 - e^{-\beta x_0}) - \bar{R}_1\, b\, \beta\, x_0} \right]. \tag{13}$$

With each trial after n_1, the conditioning proceeds with a shorter delay time $\tau' = (x_0 - \bar{x})/v$. Then in place of equation (2), we have for the intensity of the stimulus at x, when the wave front is at $\bar{x}$,

$$I(x,\bar{x}) = I_0\, e^{-\beta(x-\bar{x})}. \tag{14}$$

Since the conditioning for each trial occurs at different positions of x, or correspondingly different times t, we cannot use equation (1). Let us introduce a more general equation. If R_i is the response conditioned during the ith trial, then R_i must decrease monotonically with successive trials. Let us then consider the equation

$$R_i = F(1 - e^{-a})\, e^{-a(i-1)}\, I(x,\bar{x}). \tag{15}$$

Summing over i, we have the total response

$$R = \Sigma R_i\,. \tag{16}$$

If $I(x,\bar{x})$ is not a function of i, equation (16) with (15) reduces to (1), since

$$(1 - e^{-a}) \sum_{i=1}^{n} e^{-a(i-1)} = 1 - e^{-an}. \tag{17}$$

Substituting equation (14) into (15), and integrating according to (3), we have for the contribution to the conditioning at x by the ith trial, $(i > n_1)$

$$R(x,\bar{x})_i = F(1 - e^{-a})\, e^{-a(i-1)}$$

$$\times \left[I_0\, e^{-\beta(x-\bar{x})} - b \left\{ \int_{\bar{x}}^{x_0} R(x,n_1)\, dx + \sum_{n_1+1}^{i} R_j \right\} \right] \tag{18}$$

Now the expression in the braces is evidently the total conditioned response from all the trials and must equal $\bar{R}_1$ because of the definition of $\bar{x}$. Then introducing $\bar{R}_1$ into equation (18) and integrating with respect to x from $\bar{x}$ to x_0, we have:

$$R_T(\bar{x})_i = \frac{F}{\beta}(1 - e^{-a})\, e^{-a(i-1)} \left[I_0(1 - e^{-\beta(x_0-\bar{x})}) - \bar{R}_1\, b\, \beta(x_0 - \bar{x}) \right]. \tag{19}$$

Then summing over i from $(n_1 + 1)$ to $(n_1 + m)$ we obtain:

$$R_T(\bar{x})_m = \frac{F}{\beta}(1 - e^{-a}) \sum_{n_1+1}^{n_1+m} e^{-a(i-1)}$$

$$\times [I_0(1 - e^{-\beta(x_0 - \bar{x})}) - \bar{R}_1 b \beta (x_0 - \bar{x})] , \qquad (20)$$

where $\bar{x}$ is a function of i.

The sum of $R_T(\bar{x})_{n_1}$ and $R_T(\bar{x})_m$ is evidently the total response at $\bar{x}$ due to the conditioning from all $(n_1 + m)$ trials, and is equal to $\bar{R}_1$ from the definition of $\bar{x}$, or

$$\bar{R}_1 = R_T(\bar{x})_{n_1} + R_T(\bar{x})_m . \qquad (21)$$

If $m = m'$ for $\bar{x}(m') = x_0 - v\,\tau_0$, and if $n' = n_1 + m'$, then n' is the number of trials necessary to eliminate the alley entrance. In order to use equation (20), let us assume that $\bar{x}(i)$ in the summation has an average value which is some fraction $(1 - \eta)$ of x_0, or $x_0 - \bar{x} = \eta x_0$. Using the foregoing with equations (13) and (20), we have for equation (21)

$$\bar{R}_1 = \bar{R}_1 \left[1 - \frac{I_0(1 - e^{-\beta(x_0 - v\tau_0)}) - \bar{R}_1 b \beta (x_0 - v\,\tau_0)}{I_0(1 - e^{-\beta x_0}) - \bar{R}_1 b \beta x_0} \right]$$

$$+ \frac{F}{\beta}(1 - e^{-am'}) e^{-an_1} [I_0(1 - e^{-\beta\eta x_0}) - \bar{R}_1 b \beta \eta x_0]. \qquad (22)$$

Solving for $(1 - e^{-am'})$ we have

$$(1 - e^{-am'}) = \frac{\bar{R}_1 \beta}{F}$$

$$\times \frac{I_0(1 - e^{-\beta(x_0 - v\tau_0)}) - \bar{R}_1 \beta b (x_0 - v\,\tau_0)}{[I_0(1 - e^{-\beta x_0}) - \bar{R}_1 b \beta x_0][I_0(1 - e^{-\beta\eta x_0}) - \bar{R}_1 b \beta \eta x_0]} e^{an_1} . \qquad (23)$$

Since $n' = n_1 + m'$, we have, using equations (10) and (23),

$$n' = -\frac{1}{a}\log$$

$$\left[1 - \frac{\bar{R}_1\beta[I_0(1 - e^{-\beta(x_0 - v\tau_0)}) + I_0(1 - e^{-\beta\eta x_0}) - \bar{R}_1 b\beta(x_0 - v\tau_0 + \eta x_0)]}{F[I_0(1 - e^{-\beta x_0}) - \bar{R}_1 b\beta x_0][I_0(1 - e^{-\beta\eta x_0}) - \bar{R}_1 b\beta\eta x_0]} \right] \qquad (24)$$

which with equation (4) gives $n'(l)$.

Now equation (24) may be written

$$(1 - e^{-an'}) = (1 - e^{-an_1}) \left[1 + \frac{\varphi(x_0 - v\tau_0)}{\varphi(x_0 - \bar{x})} \right] \qquad (25)$$

where φ is defined by equation (11).

Using the first term of the expansion of the exponentials, we have for very small n's:

$$n' = n_1 \left[1 + \frac{\varphi(x_0 - v\tau_0)}{\varphi(x_0 - \bar{x})} \right]. \tag{26}$$

The following relations can be verified, as we have b very small, $\varphi > 0$, and $x_0 - v\tau_0 > \bar{x} > 0$. For $\bar{x} < v\tau_0$ or $x_0 - \bar{x} > x_0 - v\tau_0$, $\varphi(x_0 - v\tau_0)/\varphi(x_0 - \bar{x})$ increases monotonically with x_0 and, except for small values of x_0, the ratio is greater than unity. In this case $n'(l)$ is similar to the curve in Figure 1, with the minimum displaced to the left. For

$$\bar{x} > v\tau_0 \quad \text{or} \quad x_0 - \bar{x} < x_0 - v\tau_0 ,$$

$\varphi(x_0 - v\tau_0)/\varphi(x_0 - \bar{x})$ decreases monotonically with x_0, and except for small values of x_0 in a range larger than in the above case, the ratio is less than unity. Over the range for which the ratio is less than unity, n' increases with $l(x_0)$ even though the ratio itself does not. In this case, then, the curve $n'(l)$ is also similar to the curve in Figure 1, but has the minimum displaced to the right. If we take $\eta x_0 \backsim \frac{1}{2} x_0$, then the minimum of n' might be displaced to a point along x_0 which is twice as great. In either case, n' increases with l except for small values, and approaches a limit. However, if we do not restrict b to small values, $n'(l)$ may have more than one minimum, or decrease with l over a large range.

It should perhaps be emphasized that the above discussion applies not only to the elimination of a dead end but to any response pattern in which an extraneous or disagreeable secondary response is the negative of a primary response. In any case, the time τ is the time between the primary stimulus and secondary stimulus, while τ_0 represents the time after the primary stimulus during which the result of the primary response can be avoided. In certain situations, the secondary stimulus can be avoided until the very instant that it is received. In the response sequence, given in a preceding paragraph, R_a coincides with S_1 in time, and τ_0 equals τ. If we replace x_0 by $v\tau$ in equation (10), we obtain $n(v\tau)$, the number of trials to completely eliminate the extraneous or disagreeable response. The function $n(v\tau)$ is also represented by the curve in Figure 1, with the origin at the point for which $x_0 = 0$.

Since the theoretical function is obtained, it would be of interest to check with the results from an experimental determination of

$n(v\tau)$, especially if the relation is determined for very small values of τ. The existence of a lower limit for τ might then be found. It is evidently necessary to prevent anticipation of the primary stimulus. The existence of an upper limit for τ might be obscured by the formation of chain associations so that it would be necessary in the experimental procedure to minimize their effect as much as possible.

The parameters in equation (10) for $n(v\tau)$ may be obtainable from experiment. One might expect the parameters to change but slightly for various response patterns involving the same receptor and motor organs. Should the theory prove to be sufficiently complete, it would be of interest to find a possible relation of these parameters to factors obtained by other methods.

The simplified mechanism discussed here takes into account only one factor, that is, the mechanism of elimination of a wrong alley. In actual experiments, both wrong and correct alleys are present together so that there is at least one other major factor. Therefore a strict comparison of theory and experiment cannot be made at this stage. But we can show that there is no disagreement if we take the second factor into account. Let us do this here in a qualitative manner and on the basis of preliminary investigations. A more complete mathematical discussion of the problem may be presented at some later time.

The two factors may then be stated as follows: (1) Each blind alley tends to be eliminated owing to the association of the return with the entrance. It is most probable that the parameters are such that the shortest alleys would be soonest eliminated. It is this factor which has been discussed in the first section. (2) The goal response (food) becomes associated with each alley entrance, and this association is the stronger the shorter the time between the stimulus and response. Preliminary considerations show that this second factor requires (a) that the longer route to a goal response be eliminated, (b) that the alleys near the goal be eliminated more readily, (c) that the longer the alley, the sooner it be eliminated, and this holds especially for alleys near the goal.

Considering the two factors simultaneously, we would expect that the first alleys of a maze would be eliminated sooner the shorter they are, but that the order of elimination near the goal would even be inverted due to the increased importance of the second factor. The ratio of the number of trials for long alleys to the number of trials for short alleys, r, would then be a fraction greater than unity for the first alleys of the maze, and would decrease with successive alleys to a value possibly less than unity near the goal.

In order to compare these conclusions with experimental data, let us assume that the number of trials, n_1, for elimination of the alley end, is proportional to the total number of "complete entrances" made by a group of animals throughout an experiment, provided the maze is quite well learned. Additional runs then only slightly change the totals. We use n_1 here since equation (10) is simpler than equation (24). However, similar results are obtained if we use n'. Since random stimuli are present, but are not considered in the discussion, one should not be disturbed by the fact that "complete" and "partial" entrances seem to occur in an almost random order.

In the experiment by Peterson (3), two mazes were used. In Maze B, two groups of rats were used. For one group the alleys 1, 3, 5, 7, 8, 10 were shortened while for the other group the alleys 2, 4, 6, 9 were shortened. In Maze A, with six blind alleys, one group of rats ran the maze with all alleys full length, while the other group ran the maze with all alleys shortened. In order to compare ratios of the number of entrances to long alleys to that of short alleys for Maze A, all values from one group of rats were multiplied by a constant so that the grand totals of all entrances for each group would be equal. For Maze B the procedure was the same. The ratios of "long to short" entrances for each alley are shown graphically for both the A and B mazes in Figure 2. The ratios are given with the alleys equally

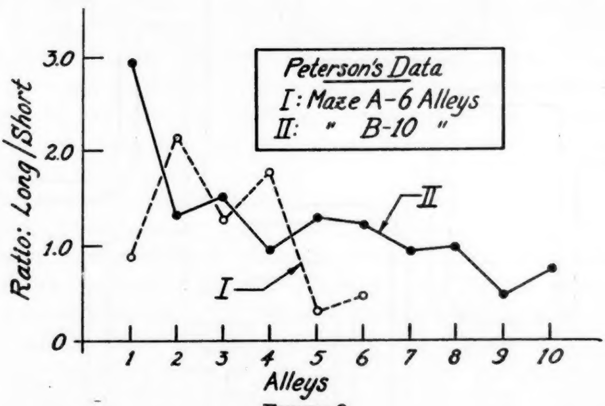


FIGURE 2

spaced, though the intervals are neither linear with distance along the maze, nor with time. In Maze A, alley 1 is the only blind alley occurring alone. The general decrease in the ratio, which occurs for both mazes, is in accord with the arguments discussed above.

In the experiments I and II by Tolman et al. (4), and the ex-

periment by White and Tolman (5), the rats were given the choice of a short blind, a long blind, and the correct path to the goal. Due to the proximity of the goal in these experiments, one would expect the second factor to predominate, so that because of factor (2c), the ratio would be less than unity. The ratios obtained, using "complete entrances," were 0.79, 1.14, and 0.47 respectively, giving a weighted average ratio of 0.82. The last group of alleys in the A and B mazes used by Peterson (5, 6 A and 8, 9, 10 B) give a weighted average ratio of 0.63.

In the case of the non-hungry rats in experiments I and II (4), there is a definite inversion, and an explanation for it is set forth. It might be added here that the strength of the association with food would not be expected to be as strong for non-hungry animals so that the relative unimportance of the second factor (2c) might well have contributed to the inversion.

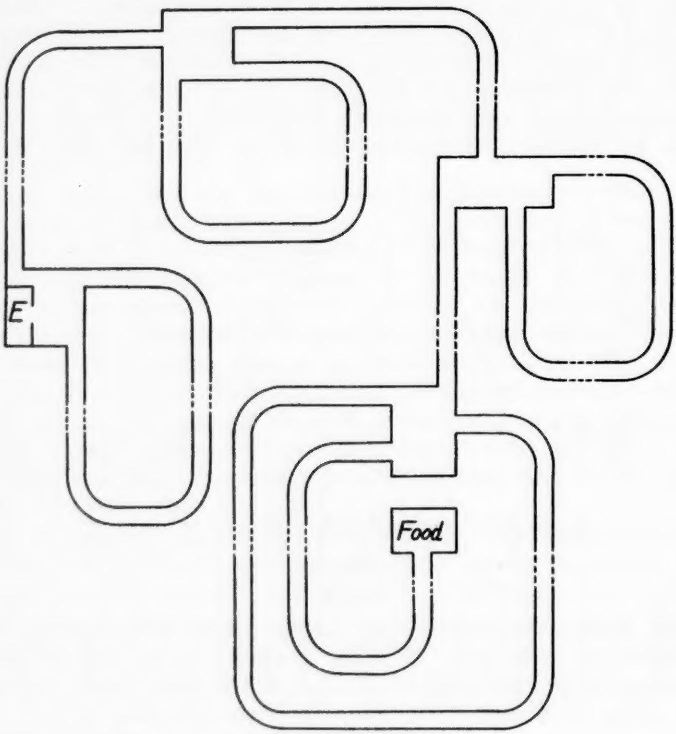


FIGURE 3

It would be of interest to use the set-up described by Tolman et al., in experiment I but with various lengths of the alley leading to food. With increased length of this path, the number of trials would be expected to increase, but there should also be an increase of the ratio of "long to short" elimination and even a possible inversion. For the length of the correct path equal to either the long or the short blind, certain peculiarities might be expected. The lengthening of the correct path might be attained by a detention cell, but here the cell might become a sufficient substitute for the goal response. Even in the case of a long alley this substitution might occur and obscure the results.

The presence of two factors, in addition to an indefinite number of small distracting factors, is already too complex. It would be desirable to eliminate one of the factors. The first factor may be minimized by the set-up shown in Figure 3, in which the animal does not have to retrace its steps. In the first unit of such a maze, one would expect the early entrances to occur according to probability. The total number of trials to obtain a fair degree of learning would decrease with successive alleys as the goal is approached, and also the ratio of "long to short" entrances would decrease from approximately unity near the entrance to a minimum near the goal, according to factor (2c).

In a maze of the type shown in Figure 3 the correct alley entrance is either to the left or right; the middle path could be made correct by raising the passageway or by suitable gates. The last section of the maze, as shown, illustrates one method of making the center path correct, and could be used for investigating the effects on maze behaviour and learning of a path which completely encircles the goal (or entrance).

We have presented a mathematical discussion of factor one, and have given a preliminary, qualitative discussion of factor two. From factor one, we expect more trials for the elimination of a longer blind alley. From factor two, we expect fewer trials for the elimination of a longer blind alley, and this holds especially near the goal. The combination of the factors gives qualitative conclusions which are born out by the negative slope of the points in the accompanying plots. There is also no disagreement in the case of other available data. Certain experimental approaches which are suggested by theoretical considerations of the problem are discussed. Indications are that the problem may be solved more readily by a somewhat more general approach. An attempt will be made to obtain this solution in the near future.

REFERENCES

1. Rashevsky, N. "Mathematical Biophysics of Delayed Reflexes in Connection with the Theory of Error Elimination." *Psychometrika*, 1936, **1**, pp. 265-273.
2. Householder, A. S., and Amelotti, E. "Some Aspects of Rashevsky's Theory of Delayed Reflexes." *Psychometrika*, 1937, **2**, pp. 255-262.
3. Peterson, Joseph. "The Effect of Length of Blind on Maze Learning." *Behaviour Monographs*, 1917, **3**, No. 4.
4. Tolman, E. C., Honzik, C. H., and Robinson, E. W. "The Effect of Degrees of Hunger upon the Order of Elimination of Short and Long Blinds." *University of California Publ. Psychol.*, 1930, **4**, pp. 189-202.
5. White, A. E., and Tolman, E. C. "A Note on the Elimination of Short and Long Blind Alleys." *Jour. Comp. Psychol.*, 1923, **3**, pp. 327-331.

PSYCHOMETRIKA—VOL. 3, NO. 3
SEPTEMBER, 1938

THE ORTHOGONAL TRANSFORMATIONS OF A FACTORIAL MATRIX INTO ITSELF

WALTER LEDERMAN
Moray House, University of Edinburgh

Certain matrix algebra, pertinent to multiple factor theory, is presented.

I. Introductory

If a team of n tests be resolved into r group factors and n specifics, the (complete) *factorial matrix* or *matrix of loadings* is an $n \times (n+r)$ matrix of the form

$$N = N_{n,\, n+r} = \begin{bmatrix} l_{11}\, l_{12} \cdots l_{1r} & m_1 & & \\ l_{21}\, l_{22} \cdots l_{2r} & & m_2 & \\ l_{n1}\, l_{n2} \cdots l_{nr} & & & m_n \end{bmatrix}, \tag{1}$$

where $l_{i\alpha}$ is the loading of the α^{th} general factor in test i and m_i is the loading of the specific in test i. On the assumption that all the factors (including the specifics) are standardized and mutually uncorrelated, the correlation matrix R and the factorial matrix N are connected by the equation

$$R = N N' . \tag{2}$$

When the tests are expressed in terms of some other set of $(n+r)$ statistically independent factors, the factorial matrix is changed to

$$\overline{N} = N B , \tag{3}$$

where B is an orthogonal matrix of order $(n+r)$. The relation (2), however, is invariant under this transformation, i.e., we have

$$R = \overline{N}\, \overline{N}' .$$

In two recent publications* Professor Godfrey H. Thomson, when discussing the indeterminacy of the factors, has constructed matrices B which not only leave the relation (2) unaltered, but also preserve the matrix of loadings itself, thus

* 1. "The Definition and Measurement of *g*," *Journal of Educational Psychology*, vol. 26, (1935), pp. 241-262.
2. "Some Points of Mathematical Technique in the Factorial Analysis of Ability," *ibidem*, vol. 27, (1936), pp. 37-54.

$$N = N B \, . \tag{4}$$

The object of this paper is *to find the general form of such a matrix, i.e., of a matrix* B *which has the following properties*:

(i) B *is orthogonal*;

(ii) B *satisfies equation* (4) .

Professor Godfrey Thomson's solution is of the form*

$$B = I - \frac{2\, q\, q'}{q'\, q}, \tag{5}$$

where

$$q = \{q_1 , q_2 , \cdots, q_r \, ; q_{r+1} , \cdots , q_{r+n}\}$$

is a column vector whose elements are defined as

$$\left. \begin{aligned} q_1 = q_2 = \cdots = q_r = -1 \, , \\ q_{r+i} = \frac{\text{sum of general loadings in test } i}{\text{specific loadings in test } i} . \end{aligned} \right\} \tag{6}$$

While in the hierarchical case ($r=1$) the matrix given by (5) and (6) is the only solution (apart from the trivial solution $B = I$), we shall see that in the case of several group factors an infinity of solutions can be found depending on $\dfrac{r(r-1)}{2}$ arbitrary parameters.

II. The General Solution

It is convenient to write the factorial matrix in the form

$$N = [L \;\vdots\; M], \tag{7}$$

where

$$L = L_{n,r} = \begin{vmatrix} l_{11} \cdots l_{1r} \\ \\ l_{n1} \cdots l_{nr} \end{vmatrix} ; M = M_{n,n} = \begin{vmatrix} m_1 & & & \\ & m_2 & & \\ & & \ddots & \\ & & & m_n \end{vmatrix} .$$

We shall always assume that M is non-singular, i.e., that none of the specifics vanishes.

Before discussing our actual problem, we shall consider a matrix Z with $(n+r)$ rows, but with an unspecified number of columns and suppose that

* See *op. cit.* (2), pp. 39-40.

$$N Z = 0 . \tag{8}$$

On writing

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix}, \tag{9}$$

where X has r rows and Y has n rows, and on using (7) we see that (8) is equivalent to

$$L X + M Y = 0 ,$$

whence

$$Y = - M^{-1} L X .$$

Substituting this in (9) we find that Z becomes

$$Z = - \begin{bmatrix} -I \\ M^{-1} L \end{bmatrix} X ,$$

or

$$Z = - Q X , \tag{10}$$

where

$$Q = \begin{bmatrix} -I \\ M^{-1} L \end{bmatrix} \tag{11}$$

is a given matrix in our problem. Hence every solution of (8) must be of the form (10) and, conversely, every matrix (10) with arbitrary X annihilates N when operating as a post-factor.

Now, if B is a solution of (4), we have

$$N(B - I) = 0 ,$$

and from the above remark it follows that

$$B - I = -Q X ,$$

or

$$B = I - Q X . \tag{12}$$

Every matrix of this form satisfies condition (ii), no matter what X is. In order to satisfy condition (i) we shall choose X in such a way that B becomes an orthogonal matrix; i.e., we must have

$$B B' = (I - Q X) (I - Q X)' = I ,$$

or

$$(I - Q X) (I - X' Q') = I . \tag{13}$$

Expanding this matrix equation and cancelling the term I on each side we obtain

$$- Q X - X' Q' + Q X X' Q' = 0 . \tag{14}$$

Before continuing with the analysis we observe that

$$Q'\,Q = I + L'\,M^{-2}\,L$$

is a positive-definite and, consequently, non-singular matrix of order r and, therefore, possesses a reciprocal and a "square root"* $(Q'\,Q)^{\frac{1}{2}}$ which is also non-singular and real.

If we now premultiply (14) by Q' and solve for the first term, we get

$$X = (Q'\,Q)^{-1}\,[Q'\,Q\,X\,X' - Q'\,X']\,Q'\,.$$

This shows that X must be of the form

$$X = T\,Q' \tag{15}$$

where T is an $(r \times r)$ matrix yet to be determined.

On substituting (15) in (12) and (13) we find

$$B = I - Q\,T\,Q' \tag{16}$$

$$(I - Q\,T\,Q')\,(I - Q\,T'\,Q') = I\,. \tag{17}$$

After premultiplying this equation by Q' and postmultiplying by Q we can write

$$(Q' - Q'\,Q\,T\,Q')\,(Q - Q\,T'\,Q'\,Q) = Q'\,Q\,,$$

$$[I - (Q'\,Q)\,T]\,(Q'\,Q)\,[I - T'\,(Q'\,Q)] = Q'\,Q\,,$$

$$\{(Q'\,Q)^{-\frac{1}{2}}\,[I - (Q'\,Q)\,T]\,(Q'\,Q)^{\frac{1}{2}}\}$$
$$\times \{(Q'\,Q)^{\frac{1}{2}}\,[I - T'\,(Q'\,Q)]\,(Q'\,Q)^{-\frac{1}{2}}\} = I\,,$$

$$\{(Q'\,Q)^{-\frac{1}{2}}\,[I - (Q'\,Q)\,T]\,(Q'\,Q^{\frac{1}{2}}\}$$
$$\times \{(Q'\,Q)^{-\frac{1}{2}}\,[I - (Q'\,Q)\,T]\,(Q'\,Q)^{\frac{1}{2}}\}' = I\,.$$

This means that

$$(Q'\,Q)^{-\frac{1}{2}}\,[I - (Q'\,Q)\,T]\,(Q'\,Q)^{\frac{1}{2}} \text{ or } I - (Q'\,Q)^{\frac{1}{2}}T\,(Q'\,Q)^{\frac{1}{2}}$$

is an r-rowed orthogonal matrix. Let

$$I - (Q'\,Q)^{\frac{1}{2}}\,T\,(Q'\,Q)^{\frac{1}{2}} = U\,.$$

Hence, solving for T,

$$T = (Q'\,Q)^{-\frac{1}{2}}(I - U)\,(Q'\,Q)^{-\frac{1}{2}}\,,$$

and, by (16),

$$B = I - Q\,(Q'\,Q)^{-\frac{1}{2}}\,(I - U)\,(Q'\,Q)^{-\frac{1}{2}}\,Q'\,. \tag{18}$$

* See M. Bôcher: *Introduction to Higher Algebra,* New York (1907), p. 299.

Conversely, it can easily be verified that any matrix B of the form (18) satisfies conditions (i) and (ii) (p. 182) provided Q is defined by (11) and U is an orthogonal matrix of order r. Thus (18) is the general solution of our problem.

It is of interest to note that different matrices U give rise to different solutions B. For suppose we had

$$I - Q(Q'Q)^{-\frac{1}{2}}(I - U_1)(Q'Q)^{-\frac{1}{2}}Q' \\ = I - Q(Q'Q)^{-\frac{1}{2}}(I - U_2)(Q'Q)^{-\frac{1}{2}}Q' ;$$

it would follow that

$$Q(Q'Q)^{-\frac{1}{2}}(I - U_1)(Q'Q)^{-\frac{1}{2}}Q' \\ = Q(Q'Q)^{-\frac{1}{2}}(I - U_2)(Q'Q)^{-\frac{1}{2}}Q' .$$

On premultiplication by Q' and postmultiplication by Q this becomes

$$(Q'Q)^{\frac{1}{2}}(I - U_1)(Q'Q)^{\frac{1}{2}} = (Q'Q)^{\frac{1}{2}}(I - U_2)(Q'Q)^{\frac{1}{2}} ,$$

and since $Q'Q$ is non-singular,

$$U_1 = U_2 .$$

Thus, there exists a (1,1) — correspondence between the solutions of the problem and all possible orthogonal matrices of order r. The general solution, therefore, depends upon $\dfrac{r(r-1)}{2}$ independent parameters.

III. Particular Solutions

1. The simplest orthogonal matrix $U = I$ always yields the trivial result

$$B = I .$$

The next simplest case is

$$U = -I .$$

This corresponds to the solution

$$B = I - 2\,Q(Q'Q)^{-\frac{1}{2}}(Q'Q)^{-\frac{1}{2}}Q' ,$$

or

$$B = I - 2\,Q(Q'Q)^{-1}Q' . \tag{19}$$

In particular, when $r = 1$, the only "orthogonal matrices" are the numbers $+1$ and -1, and the matrix Q reduces to a single column, (see eq. (11)), viz.:

$$Q = q = \{-1 , \frac{l_1}{m_1}, \frac{l_2}{m_2} \cdots , \frac{l_n}{m_n} \} ,$$

where $l_i = l_{i1}$ is the loading of the general factor in test i. The quantity $q' q$ is then a scalar and (19) becomes

$$B = I - \frac{2 q q'}{q' q}.$$

This is the formula at which Professor Godfrey Thomson arrived in the first of the two papers cited on p. 181.

2. If

$$e = \{e_1, e_2, \cdots\}$$

is any non-zero column vector, it is easy to prove that

$$U = I - \frac{2 e e'}{e' e} \tag{20}$$

is an orthogonal matrix, i.e., that

$$U U' = I.$$

Substituting (20) in (18) we obtain

$$B = I - Q (Q' Q)^{-\frac{1}{2}} \left(\frac{2 e e'}{e' e}\right) (Q' Q)^{-\frac{1}{2}} Q'.$$

Now put

$$(Q' Q)^{-\frac{1}{2}} e = a,$$

or

$$e = (Q' Q)^{\frac{1}{2}} a,$$

where

$$a = \{a_1, a_2, \cdots, a_r\}$$

is an arbitrary column vector of order r. We then get

$$B = I - 2 \frac{(Qa)(Qa)'}{(Qa)'(Qa)}.$$

In particular, if

$$a = \{1, 1, \cdots, 1\},$$

we find that

$$Q\,a = \begin{bmatrix} -1 & & & \\ & -1 & & \\ & & \ddots & \\ & & & -1 \\ \frac{l_{11}}{m_1} & \frac{l_{12}}{m_1} & \cdots & \frac{l_{1r}}{m_1} \\ \frac{l_{21}}{m_2} & \frac{l_{22}}{m_2} & \cdots & \frac{l_{2r}}{m_2} \\ \frac{l_{n1}}{m_n} & \frac{l_{n2}}{m_n} & \cdots & \frac{l_{nr}}{m_n} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} =$$

$$\{-1\,, -1\,, \cdots, -1\,; \frac{\sum l_{1a}}{m_1}, \frac{\sum l_{2a}}{m_2}, \cdots, \frac{\sum l_{na}}{m_n}\}\,.$$

The corresponding solution can be written in the form

$$B = I - 2\,\frac{q\,q'}{q'\,q}\,,$$

where

$$q = \{\,q_1\,, q_2\,, \cdots, q_r\,;\, q_{r+1}\,, \cdots, q_n\}$$

and

$$q_1 = q_2 = \cdots = q_r = -1\,,$$
$$q_{r+i} = \frac{\sum l_{ia}}{m_i}\,,$$

which is identical with Professor Godfrey Thomson's results (5) and (6).

PSYCHOMETRIKA—VOL. 3, NO. 3
SEPTEMBER, 1938

A METHOD FOR FINDING THE INVERSE OF A MATRIX

LEDYARD R. TUCKER
The University of Chicago

The problem of solving simultaneous linear equations is one that frequently confronts the scientist. Whenever it is desired to state the parameters as expressions of the constant terms of these equations, the problem resolves itself into the discovery of the inverse of the matrix of coefficients. The method here described gives a routine for the solution of this problem.

Consider the following set of simultaneous linear equations:

$$\begin{aligned} .80\, b_1 + .48\, b_2 + .36\, b_3 &= d_1 , \\ .48\, b_1 + .80\, b_2 + .36\, b_3 &= d_2 , \\ .36\, b_1 + .36\, b_2 + .86\, b_3 &= d_3 . \end{aligned} \tag{1}$$

These equations may also be written in matrix form:

$$\underset{A}{\begin{Vmatrix} .80 & .48 & .36 \\ .48 & .80 & .36 \\ .36 & .36 & .86 \end{Vmatrix}} \cdot \underset{B}{\begin{Vmatrix} b_1 \\ b_2 \\ b_3 \end{Vmatrix}} = \underset{D}{\begin{Vmatrix} d_1 \\ d_2 \\ d_3 \end{Vmatrix}} . \tag{1m}$$

The problem is to state the "b's" in terms of the "d's."

This problem may be reduced to a new form by defining two new matrices. Consider the following matrix equation:

$$\underset{A^{-1}}{\begin{Vmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{Vmatrix}} \cdot \underset{A}{\begin{Vmatrix} .80 & .48 & .36 \\ .48 & .80 & .36 \\ .36 & .36 & .86 \end{Vmatrix}} = \underset{I}{\begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}} . \tag{2m}$$

The matrix with unity in the diagonal cells and zeros in all other cells is known as the identity matrix. The matrix A^{-1} is that matrix which, when multiplied by the matrix A, gives the identity matrix as the product. This definition of A^{-1} is implied in equation (2m). This matrix A^{-1} is known as the inverse of matrix A. Both members of equation (1m) may be premultiplied by the matrix A^{-1}:

$$\underset{A^{-1}}{\begin{Vmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{Vmatrix}} \cdot \underset{A}{\begin{Vmatrix} .80 & .48 & .36 \\ .48 & .80 & .36 \\ .36 & .36 & .86 \end{Vmatrix}} \cdot \underset{B}{\begin{Vmatrix} b_1 \\ b_2 \\ b_3 \end{Vmatrix}} = \underset{A^{-1}}{\begin{Vmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{Vmatrix}} \cdot \underset{D}{\begin{Vmatrix} d_1 \\ d_2 \\ d_3 \end{Vmatrix}} .$$

This equation can be simplified by the use of equation (2m) to the following form:

$$\underset{I}{\begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}} \cdot \underset{B}{\begin{Vmatrix} b_1 \\ b_2 \\ b_3 \end{Vmatrix}} = \underset{A^{-1}}{\begin{Vmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{Vmatrix}} \cdot \underset{D}{\begin{Vmatrix} d_1 \\ d_2 \\ d_3 \end{Vmatrix}} .$$

The important property of the identity matrix is that the product of it and a matrix B is the matrix B. For this reason it has been denoted the "identity matrix." Using this property, the above equation becomes:

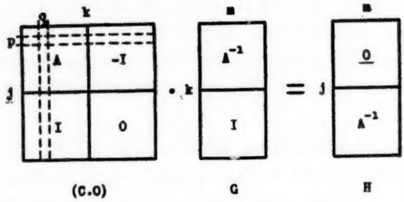
$$\underset{B}{\begin{Vmatrix} b_1 \\ b_2 \\ b_3 \end{Vmatrix}} = \underset{A^{-1}}{\begin{Vmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{Vmatrix}} \cdot \underset{D}{\begin{Vmatrix} d_1 \\ d_2 \\ d_3 \end{Vmatrix}} . \tag{3m}$$

Thus it is seen that the inverse of a matrix in matrix algebra corresponds to the reciprocal of a number in ordinary algebra. The problem is now to find the matrix A^{-1}. A method for accomplishing this, originally developed by A. C. Aitken[1], here is simplified and provided with complete numerical checks.

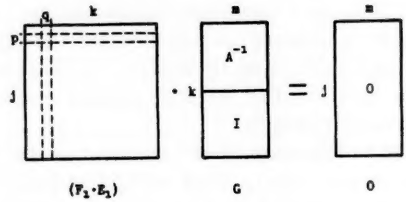
The first step in the solution is to build up the matrix $(C.0)$, as shown in equation (4m). The matrix A is copied in the upper left section of $(C.0)$.The upper right section has minus one's (—1) in the cells of the principal diagonal and zeros (0) in all other cells, thus forming a negative identity matrix. The lower left section is a positive identity matrix with plus one's (+1) in the principal diagonal. The lower right section is completely filled with zeros (0). All four sections are of the same order.

Continuing the construction of equation (4m): the matrix G has A^{-1} in the upper section and an identity matrix in the lower section.

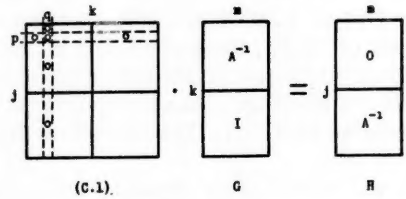
[1] Thomson, Godfrey H. "Some Points of Mathematical Technique in the Factorial analysis of Ability," *Journal of Educational Psychology*, January, 1936, **27**, 37.


q
k
m
m
p
A
-I
A^{-1}
0
j
k
j
I
0
I
A^{-1}
(C.0)
G
H

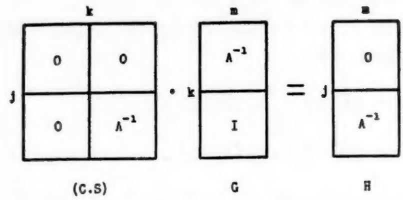
(4m)


q
k
m
m
p
A^{-1}
j
k
j
0
I
($F_1 \cdot E_1$)
G
0

(5m)


q
k
m
m
p
A^{-1}
0
j
k
j
I
A^{-1}
(C.1)
G
H

(6m)


k
m
m
0
0
A^{-1}
0
j
k
j
0
A^{-1}
I
A^{-1}
(C.S)
G
H

(7m)

The matrix H has zeros (0) in the upper half and the matrix A^{-1} in the lower section. These matrices, G and H, may be indicated in equation (4m) even though their entries are unknown. The equality of equation (4m) may be verified by performing the indicated multiplication; this is done in equation (8m). This multiplication makes use of equation (2m).

The final equation in the solution is equation (7m). Matrix $(C.S)$ has zeros (0) in all cells except those in the lower right section, this section being occupied by A^{-1}. That this equation is also an equality may be readily seen by performing the indicated multiplication. It will be noted that to find A^{-1} it is only necessary to transform matrix $(C.0)$ to matrix $(C.S)$ in such a manner as not to alter the equality to matrix H when $(C.S)$ is postmultiplied by matrix G. This can be accomplished by a procedure composed of several similar steps. The first of these steps will be traced through.

The first important point in this step of the solution is the definition of matrix $(F_1.E_1)$ of equation (5m). This matrix is such a matrix that, when postmultiplied by matrix G, a matrix filled with zeros is produced. This matrix is further defined to have a row equal to one of the rows in the upper half of $(C.0)$ and a column equal to one of the columns in the left half of $(C.0)$, these rows and columns occupying corresponding positions in the two matrices. This condition is shown in equation (5m), in which $(F_1.E_1)$ has row p and column q equal to the same row and column of $(C.0)$. The procedure for finding this matrix will be shown later.

When equation (5m) is subtracted from equation (4m), the result is equation (6m). In this subtraction, matrix G can be factored out, and matrix H has a zero matrix subtracted from it; therefore, neither is altered. The matrix $(C.1)$ is defined by equation (9m) and has entries as given in equation (9).

$$(C.1) = (C.0) - (F_1.E_1) . \tag{9m}$$

$$(C.1)_{jk} = (C.0)_{jk} - (f_j.e_k) . \tag{9}$$

Since the entries in row p and column q of $(F_1.E_1)$ are equal to the entries in the same row and column of $(C.0)$, the entries in this row and column of $(C.1)$ are all zeros. The matrix (C.1) is, therefore, one step nearer $(C.S)$ than is $(C.0)$. In order to complete the solution it is only necessary to repeat the procedure outlined above. Obtain the matrix $(C.2)$ from matrix $(C.1)$ in the same manner that $(C.1)$ was obtained from $(C.0)$, and then proceed from $(C.2)$ in the same way, until a matrix with the whole upper half and left half filled with

$$\begin{array}{c} m \\ j\begin{bmatrix} A\cdot A^{-1} - I\cdot I \\ \hline I\cdot A^{-1} + 0 \end{bmatrix} \\ (C.O)\cdot G \end{array} = \begin{array}{c} m \\ j\begin{bmatrix} I - I \\ \hline A^{-1} \end{bmatrix} \\ (C.O)\cdot G \end{array} = \begin{array}{c} m \\ j\begin{bmatrix} 0 \\ \hline A^{-1} \end{bmatrix} \\ H \end{array} \tag{8m}$$

$$\begin{array}{c} k \\ \begin{bmatrix} e_k \end{bmatrix} \\ E_1 \end{array} \cdot \begin{array}{c} m \\ k\begin{bmatrix} A^{-1} \\ \hline I \end{bmatrix} \\ G \end{array} = \begin{array}{c} m \\ \begin{bmatrix} 0 \end{bmatrix} \\ 0 \end{array} \tag{10m}$$

$$e_k = (C.O)_{pk} / (C.O)_{pq} \tag{10}$$

$$\begin{array}{c} \\ j\begin{bmatrix} f_j \end{bmatrix} \\ F_1 \end{array} \cdot \begin{array}{c} k \\ \begin{bmatrix} e_k \end{bmatrix} \\ E_1 \end{array} \cdot \begin{array}{c} m \\ k\begin{bmatrix} A^{-1} \\ \hline I \end{bmatrix} \\ G \end{array} = \begin{array}{c} m \\ j\begin{bmatrix} 0 \end{bmatrix} \\ 0 \end{array} \tag{11m}$$

$$f_j = (C.O)_{jq} \tag{11}$$

zeros is obtained. This matrix is the matrix $(C.S)$ and contains the inverse of matrix A as the lower right quarter.

In order to build the matrix $(F_1.E_1)$, the matrices F_1 and E_1 must first be defined. The matrix E_1, in equation (10m), is proportional to row p of matrix $(C.0)$, the constant of proportionality being the reciprocal of the entry in the row p and column q of matrix $(C.0)$. Equation (10) shows the value of the k^{th} entry of E_1. Since, in matrix $(C.0)$ of equation (4m), this row, p, times the matrix G is equal to a row of zeros in H, equation (10m) is an equality. In equation (11m) both sides of equation (10m) are premultiplied by a column matrix F_1, the entries in F_1 being equal to the entries in the column q of $(C.0)$, as is shown in equation (11). The matrix on the right side of the equation is obviously filled with zeros, for it is the product of two matrices, one of which has zeros in all cells. The matrix $(F_1.E_1)$ is the product of matrices F_1 and E_1. In equations (12) and (13) it is shown that the entries in row p and column q of matrix $(F_1.E_1)$ are equal to the entries in the same row and column of matrix $(C.0)$.

$$f_p.e_k = (C.0)_{pq}.(C.0)_{pk}.1/(C.0)_{pq},$$

$$f_p.e_k = (C.0)_{pk} = (F_1.E_1)_{pk}. \tag{12}$$

Similarly:

$$f_j.e_q = (C.0)_{jq} = (F_1.E_1)_{jq}. \tag{13}$$

In the actual calculational procedure the matrix $(C.0)$ is built up as has been described; that is, it has the matrix A in the upper left quarter, minus one's in the diagonal cells of the upper right quarter, plus one's in the diagonal cells of the lower left quarter, and zeros in all other cells. The matrix $(C.0)$ for the illustrative problem is shown in Table 1. The selection of row p and column q depends, in practice, upon the selection of the entry included in both of them. Several factors enter into the selection of the pivot entry. The first of these considerations is that it must be located in the upper left section of $(C.0)$ in order to satisfy the conditions already developed in the theory of the solution. The second factor to be considered in the selection of the pivot entry is that it should be quite large numerically in order to minimize computational error. In the illustrative problem in Table 1 the entry in the first row and first column of $(C.0)$ was selected as the pivot entry. This entry is in the upper left section of $(C.0)$ and is the second largest entry in this section.

After the pivot entry has been selected, the next step in the calculation of the inverse is to find the row matrix E_1. The reciprocal of the pivot entry is found first; for example, in the illustrative problem:

$$1/.800 = 1.250000 .$$

An entry in the row E_1 is equal to the product of the entry in the same column of row p and the reciprocal of the pivot entry. The formula for this calculation is given in equation (10). For an example in the illustrative problem the second entry in E_1 is:

$$.480 \times 1.250000 = .600 .$$

where .480 is the second entry in the first, or p^{th}, row of $(C.0)$. A check on these calculations can be obtained by multiplying the sum of row p by the reciprocal of the pivot entry; this product should equal the sum of the row E_1. The formula for this computational check is:

$$\sum_k e_k = \left[\sum_k (C.0)_{pk}\right] 1/(C.0)_{pq} .$$

The column F_1 is secured by copying the column q of $(C.0)$. In the illustrative problem the column F_1 is a copy of the first, or q^{th}, column of $(C.0)$.

The next step in the solution is to find the matrix $(C.1)$. The recording of the matrix $(F_1.E_1)$ is omitted in practice, the procedure being to combine its calculation with the subtraction from $(C.0)$. The equation for each entry in $(C.1)$ is given by equation (9). This equation states that the entry in row j and column k of $(C.1)$ is equal to the corresponding entry in $(C.0)$ minus the product of the j^{th} entry in F_1 and the k^{th} entry in E_1. In the illustrative problem the entry in row II and column C of $(C.1)$ is:

$$.360 - (.480 \times .450) = .144 .$$

The .360 is found in the row II and column C of $(C.0)$. The entry in row II of F_1 is .480, and the entry in the column C of E_1 is .450. A check of these calculations can be obtained by treating the sums of the columns of $(C.0)$ in the same manner as the individual entries of $(C.0)$. The entry in the column F_1 for this calculation is the sum of the column F_1. The result of this computation should equal the sum of the corresponding column in $(C.1)$. The formula for the calculation of this check is:

$$\sum_j (C.1)_{jk} = \sum_j (C.0)_{jk} - e_k \sum_j f_j .$$

In the illustrative problem the check on the sum of column C in $(C.1)$ is:

$$2.580 - (.450 \times 2.640) = 1.392 .$$

When the entries in $(C.1)$ have been computed, the next portion of the solution is a duplicate of the foregoing steps. A pivot entry is

selected and the row E_2 and column F_2 found. The entries in the matrix $(C.2)$ are then found from the entries in $(C.1)$, E_2, F_2 in the same manner that the entries in (C.1) were computed from the entries in $(C.0)$, E_1, and F_1. The selection of the pivot entry in $(C.1)$ has the same limitations as the selection of the pivot entry in $(C.0)$, namely: the pivot entry must be in the upper left section, and it should be one of the largest entries in this section in absolute value.

The foregoing process is repeated until all the entries in the upper half, and in the left half of the final matrix are zeros. The inverse of the matrix A is then in the lower right section of this final matrix, $(C.S)$. This condition is shown in equation (7m) and in matrix $(C.3)$ of the illustrative problem. A final check of the calculations may be made by substituting A^{-1} into equation (2m); for the illustrative problem, that is:

$$\underset{A^{-1}}{\begin{Vmatrix} 2.073 & -1.052 & -.427 \\ -1.052 & 2.073 & -.427 \\ -.427 & -.427 & 1.520 \end{Vmatrix}} \cdot \underset{A}{\begin{Vmatrix} .80 & .48 & .36 \\ .48 & .80 & .36 \\ .36 & .36 & .86 \end{Vmatrix}} = \underset{I}{\begin{Vmatrix} 1.000 & .000 & .000 \\ .000 & 1.000 & .000 \\ .001 & .001 & 1.000 \end{Vmatrix}} .$$

The calculation of the inverse of matrix A is then complete.

(C.0)

	A	B	C	I	II	III	Σ	F_1
I	.800	.480	.360	-1.000	.000	.000	.640	.800
II	.480	.800	.360	.000	-1.000	.000		.480
III	.360	.360	.860	.000	.000	-1.000		.360
A	1.000	.000	.000	.000	.000	.000		1.000
B	.000	1.000	.000	.000	.000	.000		.000
C	.000	.000	1.000	.000	.000	.000		.000
Σ	2.640	2.640	2.580	-1.000	-1.000	-1.000	4.860	2.640
E_1	1.000	.600	.450	-1.250	.000	.000	.800 = Ch.ΣE_1	
	1/.800 = 1.250000						.800 = ΣE_1	

(C.1)

	A	B	C	I	II	III	Σ	F_2
I								
II		.512	.144	.600	-1.000	.000	.256	.512
III		.144	.698	.450	.000	-1.000		.144
A		- .600	- .450	1.250	.000	.000		- .600
B		1.000	.000	.000	.000	.000		1.000
C		.000	1.000	.000	.000	.000		.000
Ch.		1.056	1.392	2.300	-1.000	-1.000	2.748	
Σ		1.056	1.392	2.300	-1.000	-1.000	2.748	1.056
E_2		1.000	.281	1.172	-1.953	.000	.500 = Ch.ΣE_2	
	1/.512 = 1.953125						.500 = ΣE_2	

(C.2)

	A	B	C	I	II	III	Σ	F_3
I								
II								
III			.658	.281	.281	-1.000	.220	.658
A			-.281	1.953	-1.172	.000		-.281
B			-.281	-1.172	1.953	.000		-.281
C			1.000	.000	.000	.000		1.000
Ch.			1.095	1.062	1.062	-1.000	2.220	
Σ			1.096	1.062	1.062	-1.000	2.220	1.096
E_3			1.000	.427	.427	-1.520	.334 = Ch.ΣE_3	
	1/.658 = 1.519757						.334 = ΣE_3	

(C.3)

	A	B	C	I	II	III	Σ	
I								
II								
III								
A				2.073	-1.052	-.427		
B				-1.052	2.073	-.427		
C				-.427	-.427	1.520		
Ch.				.594	.594	.666	1.854	
Σ				.594	.594	.666	1.854	

TABLE I

ANNOUNCEMENT

A manuscript by Professor N. Rashevsky entitled "Contribution to the Mathematical Biophysics of Visual Perception with Special Reference to the Theory of Aesthetic Values of Geometrical Patterns" was scheduled for this issue of Psychometrika. Owing to an editorial delay, publication of this article has been postponed to the December issue.